

ESSAYS IN CREDIT DERIVATIVES

BY

CHANATIP KITWIWATTANACHAI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Finance
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Neil D. Pearson, Chair
Professor Heitor Almeida, Director of Research
Professor Timothy Johnson
Professor George Pennacchi

Abstract

This thesis consists of three essays that examine various problems in credit derivatives. In the first essay, we propose a novel method to extract asset correlations from credit derivatives. Default correlation is a concern especially after witnessing the financial crisis. To find default correlations, we would like to know asset correlations which are unobservable. We derive a model to infer asset correlations from Credit Default Swaps (CDS). We use a structural model approach with the first passage time as default. The resulting model is closed-form and extremely easy to compute. Using the data from 2004 to 2008, we find the average implied asset correlation from CDS to be over 0.4. The average equity correlation, which is usually used as a proxy for asset correlation, over the same period is 0.155. The result complies with the literature that there is another unobservable factor driving defaults among firms.

The second essay examines the illiquidity of the CDS market. Researchers claim that CDS spreads reflect "purer" default risk than the bond spreads. We investigate whether the CDS market is really liquid. Since it is hard to define and measure liquidity precisely, we use an event study to answer the question. The event is when a CDS is included into the CDX index. This event changes the liquidity of CDS because they will be traded in the more liquid CDX market. If the CDS market is already liquid, we should observe no change in the level or correlation of CDS spreads. However, the spread levels do change, and the correlations between CDS and CDX index also change, to a lesser extent. The significant changes in the spread levels suggest that the CDS market is not perfectly liquid. The most likely channel for illiquidity is that order imbalance causes price impact depending on the direction of dealers' inventory.

The third essay shows that stochastic recovery rates are priced ex ante in CDS and thus we can extract this information from CDS spreads. Recovery rates have been treated as a constant in the literature. However, recent empirical findings suggest that realized recovery rates are also stochastic and highly dependent on the industry condition. It is particularly hard to separate the effect of risk-neutral probability of default and risk-neutral recovery rates in CDS. We use the unique characteristic of ex post (physical) recovery rates to capture the ex ante (risk-neutral) recovery rates in CDS spreads. We find that the stochastic recovery rates affect the CDS spreads, ex ante. If the industry is in distress, the risk-neutral recovery rates are expected to be lower. We derive a simple first-passage-time structural model to capture the empirical findings.

If the industry is in distress, the expected risk-neutral recovery rate will be lower by 20%. The model can be used to learn about expected recovery rates across business cycles from CDS data.

To my Father and Mother.

Acknowledgments

I am indebted to my advisor, Neil Pearson, for his guidance and wisdom. I thank the members of my dissertation committee: Tim Johnson, George Pennacchi and Heitor Almeida. I would also like to thank all other members of the Department of Finance at the University of Illinois for their assistance and mentoring. Thanks to my mother and my farther for encouraging the pursuit of higher education. Thanks to the King's scholarship from Thailand for me to start my journey in the US. Financial support from the Irwin Fellowship is gratefully acknowledged.

Table of Contents

Chapter 1: Inferring Asset Correlations from CDS Spreads.....	1
1.1 Introduction	1
1.2 Model.....	5
1.3 Empirical Analysis.....	10
1.4 Robustness Check.....	11
1.5 Default Simulation.....	14
1.6 Interpretation.....	16
1.7 Conclusion and Future Work.....	17
1.8 Figures and Tables.....	18
Chapter 2: The Illiquidity of CDS Market.....	27
2.1 Introduction.....	27
2.2 CDX Index and Inclusion and Exclusion Procedure.....	30
2.3 Data.....	33
2.4 Regression Analysis and Event Study.....	35
2.5 Robustness Check.....	42
2.6 Interpretation.....	48
2.7 Conclusion and Future Work.....	50
2.8 Figures and Tables.....	52
Chapter 3: The Recovery Factor in Credit Default Swaps.....	63
3.1 Introduction.....	63
3.2 CDS Pricing Model.....	70
3.3 The Probability of Default.....	73
3.4 Data.....	78
3.5 Empirical Results.....	80
3.6 Robustness Check.....	83
3.7 Modified Theoretical Model.....	94
3.8 Conclusion and Future Work.....	105
3.9 Figures and Tables	107

Appendix A: Proofs.....	127
Bibliography.....	139

Chapter 1

Inferring Asset Correlations from CDS Spreads

1.1 Introduction

Credit risk arises from the possibilities that the underlying assets in transactions may default. Credit risk has a major impact on the valuation of financial products such as the Collateralized Debt Obligations (CDOs) and portfolio credit risk. A mistake in assessing credit risk may result in significant losses as can be seen from the recent financial crisis. To quantify credit risk, we want to know the probability of default of each firm and joint probability of default among firms. Unfortunately, defaults are rare, thus making it hard to build a model to assess default probabilities. Even harder to study are default correlations. This is because asset correlations, a crucial input for the model, are unobservable. Some industry practitioners have tried to use equity correlations to predict default correlations but found the result

unsatisfactory.

There are two main approaches to model credit risk: structural models and reduced-form models. A structural model characterizes defaults as asset values falling below default barriers. A reduced-form model abstracts away from the fundamentals of the firm and characterizes defaults as a Poisson process. The advantage of a structural model is that it incorporates the dynamics of asset values and default barriers, thus providing an economically-meaningful dynamic model. Unfortunately, asset values and default barriers are unobservable. Moreover, using the first passage time as default (Black and Cox, 1976), the model quickly becomes mathematically and computationally hard to solve.

Despite the mathematical difficulties, the intuitive and dynamical aspects of the structural model make it appealing to model default correlations. First, though, we have to overcome the problem that asset values and default barriers are unobservable. Fortunately, with the more recent market of Credit Default Swaps (CDSs), we have a new dataset that should reflect asset values and default barriers. Essentially, a CDS is an insurance against default risk, and thus CDS spreads should reflect the probability of default of the underlying asset. Moreover, the time series of CDS spreads should reflect the dynamics of asset values, from which we can find asset correlations. The link between CDS spreads and firm's fundamentals has to be there, somewhere.

In this paper we establish that link. We develop a theoretical model to infer asset correlations from CDS spreads. We derive a closed-form solution for the probability of default and asset correlations in terms of CDS spreads. The model is computationally efficient, requiring only a numerical integra-

tion and differentiation. The resulting asset correlations are strikingly higher than the corresponding equity correlations over the same period. This result is in fact consistent with the literature which found that there is an extra unobservable factor, apart from the well-known financial variables, that drives defaults among firms (see, e.g., Collin-Dufresne, Goldstein and Martin (2001) and Das et al. (2007)).

Merton (1974) pioneers a structural model which relates the unobservable firm value to the equity value. Black and Cox (1976) extend the model to characterize defaults as the first passage time of the firm value across a barrier. The first passage time distribution of a Brownian motion across a barrier can be found in the books by Harrison (1985) or Shreve (2004). Zhou (2001) derives the probability of default when the barrier grows at the same rate as the expected growth rate of the firm. However, he assumes that asset correlations and distance-to-default – a crucial input for the model, generally unobservable – are known. This paper fills the gap by deriving a model to find asset correlations and distance-to-default from CDS data.

Hull, Predescu and White (2009) also use a structural model approach with first passage time to extract the mean asset correlation from CDO data. However, they do not derive the solution in an explicit form and solve the model instead by a Monte-Carlo simulation. Also, their approach cannot find pairwise asset correlations, only the mean. Our model finds the probability of default in an explicit form. Moreover, to the best of our knowledge, our model is the first to derive the dynamics of CDS spreads in terms of the underlying asset's Brownian motion in closed-form. As such, we can readily find asset correlations directly from CDS spreads without the need of Monte-

Carlo simulation. We can find pairwise asset correlations given CDS data of any two firms. Giesecke (2004) provides a model for correlated default with incomplete information, also based on a structural model. The key to default correlation in his case is that investors suddenly learn about the barriers of other firms associated to the defaulted firm. Our paper is concerned with extracting the asset correlation parameter from credit derivatives instruments. There is no surprise in the firm's value or barrier in our model.

The chapter proceeds as follows. In Section 1.2, we extend a structural model to address the relationship between CDS spreads and the probability of default and asset dynamics. To make the model as simple as possible, we need a "nice" assumption, which is, the barrier growing at an exponentially deterministic rate equal to the asset's expected growth rate. In Section 1.3, we use the theoretical result to extract the probability of default and asset correlations from CDS data from the period of 2004 to 2008, and compare them to equity correlations. In Section 1.4, we relax the "nice" assumption and incorporate more realistic assumptions from recent literature. We show that the results still remain the same. In Section 1.5, we simulate an artificial economy using the first passage time framework to find the true implied correlations. In Section 2.6, we interpret the difference between implied correlations from credit and equity markets. Finally we end with conclusion and future work.

1.2 Model

A Credit Default Swap (CDS) is a contract that provides insurance against the risk of a default by a particular company. The company is known as the reference entity and a default by the company is known as a credit event. The buyer of this contract obtains the right to sell bonds issued by the reference entity for their face value when a credit event occurs. On the other hand, the CDS seller agrees to buy the bonds for their face value when a credit event occurs, thus bearing the risk of default.

The CDS buyer makes periodic payments to the seller until the end of the life of the CDS or until a credit event occurs. The settlement, in the event of a default, involves either physical delivery of the bonds or a cash payment. The observable quantity in the market is the payment by the CDS buyer, also known as CDS spreads. The CDS spread in a liquid market reflects the fair price of default insurance, i.e., the spread must make the expected value of the buyer's periodic payments equal to the expected value of the seller's losses in case of a default .

We take CDS spread formula, given for the contract starting from time 0 to T , from Hull, Predescu and White (2009):

$$S = \frac{(1 - \hat{R})(1 + a^*) \int_0^T q(u)v(u)du}{\int_0^T q(u)[h(u) + e(u)]du + (1 - \int_0^T q(u)du)h(T)} \quad (1.1)$$

where

T : The period of the contract

$q(u)$: Probability density function (pdf) of default at time u in a risk-neutral world

\hat{R} : Expected recovery rate on the reference obligation in a risk-neutral world

$h(u)$: Present value of payments at the rate \$1 per year on payment dates between time zero and time u

$e(u)$: Present value of accrual payment at time u equal to $u - u^*$ where u^* is the payment date immediately preceding time u

$v(u)$: Present value of \$1 received at time u

a^* : Average value of accrued interest rate on the reference obligation for the period 0 to T

The important quantity in this equation is $q(u)$. In a structural model we characterize default as the first passage time of the firm value (V) across the barrier (B). We assume the barrier grows at the same rate as the expected growth rate of the firm, i.e., the firm has a constant expected leverage ratio in the risk-neutral measure. We choose this assumption mainly to find the simplest possible model to express CDS spread dynamics in terms of asset dynamics. This assumption turns out to simplify much of the mathematics involved. Economically, this assumption is supported by recent papers. Almeida and Philippon (2007) argue that firms evaluate capital structure decisions in the risk-neutral measure. Collin-Dufresne and Goldstein (2001) find that a structural model with mean-reverting leverage ratios is more consistent with empirical findings. Considering the two papers together, we assume in the model that firms maintain a constant expected leverage ratio in the risk-neutral measure. Later on we will relax this assumption and implement the more realistic mean-reverting model as in Collin-Dufresne and Goldstein (2001) and confirm that the results do not change.

The probability of default is equal to the first passage time distribution.

The analytical formula is well-known as given by Harrison (1985) and Shreve (2004) and is also used by Zhou (2001). We summarize the basic setup and the result here while the proof can be found in Appendix A.1.

In the risk-neutral measure, we have

$$\frac{dV}{V} = rdu + \sigma dW^Q$$

Let $B(u)$ be the default barrier at time u . By our assumption,

$$B(u) = B(0)e^{(r - \frac{\sigma^2}{2})u}$$

With this setup, we get the default density (or first passage time density) in the risk-neutral measure:

$$q(m, u) = \frac{m}{u\sqrt{2\pi u}} e^{-\frac{m^2}{2u}}$$

where $m = \frac{\log(\frac{V(0)}{B(0)})}{\sigma}$. Note that m is equivalent to the distance-to-default at the beginning of CDS contract. The density $q(m, u)$ here is indeed what we want for $q(u)$ in equation 3.2.

Figure 1.1 shows the probability density of default, $q(m, u)$, for selected values of m corresponding to different credit ratings. Figure 1.2 shows the historical default density from S&P report for firms with different credit ratings. Note that the two figures show similar patterns of defaults. We conclude that the theoretical formula of default density, $q(m, u)$, can capture real-world default probabilities, and thus can be used to calculate CDS spreads.

Now we consider the CDS spread at any contract time t to $t + T$. We write (3.2) as follows:

$$S(m, t) = \frac{(1 - \hat{R})(1 + a^*) \int_t^{t+T} q(m, u - t) v(u - t) du}{\int_t^{t+T} q(m, u - t) [h(u - t) + e(u - t)] du + (1 - \int_t^{t+T} q(m, u - t) du) h(t + T)}$$

where $m = \frac{\log(\frac{V(t)}{B(t)})}{\sigma}$, the distance-to-default at the beginning of the contract at time t . In fact the integration limits of this expression do not depend on t . Indeed, let $w = u - t$, we get

$$\begin{aligned} S(m) &= \frac{e^{-rt} \left[(1 - \hat{R})(1 + a^*) \int_0^T q(m, w) v(w) dw \right]}{e^{-rt} \left[\int_0^T q(m, w) [h(w) + e(w)] dw + (1 - \int_0^T q(m, w) dw) h(T) \right]} \\ &= \frac{(1 - \hat{R})(1 + a^*) \int_0^T q(m, w) v(w) dw}{\int_0^T q(m, w) [h(w) + e(w)] dw + (1 - \int_0^T q(m, w) dw) h(T)} \end{aligned} \quad (1.2)$$

In other words, CDS spreads depend only on the distance-to-default at the beginning of the contract. CDS spreads as a function of distance-to-default (m) is shown in Figure 1.3. We numerically solve (1.2) using $T = 5$, risk-free rate (r) = 0.025, recovery rate (\hat{R}) = 0.4 and $a^* = 0.004$.

The figure is self-explanatory and intuitive. Lower distance-to-default corresponds to higher probability of default and higher CDS spreads, and vice-versa. From this graph, we can infer the distance-to-default (m) from observable CDS spreads and calculate the risk-neutral probability of default before time u using the formula (proof in Appendix A.1):

$$P \{ \tau_m \leq u \} = 2\Phi\left(\frac{-m}{\sqrt{u}}\right)$$

where τ_m is the time of default depending on m . Now we want to find the dynamics of CDS spreads as a function of the underlying asset's dynamics.

We write the dynamics of $S(m)$ using Ito's lemma:

$$dS = \frac{\partial S}{\partial m} dm + \frac{1}{2} \left(\frac{\partial^2 S}{\partial m^2} \right) (dm)^2 \quad (1.3)$$

Now consider $m(V, B)$. Using Ito's lemma, we obtain:

$$\begin{aligned} dm &= \frac{\partial m}{\partial V} dV + \frac{1}{2} \left(\frac{\partial^2 m}{\partial V^2} \right) (dV)^2 + \frac{\partial m}{\partial B} dB \\ &= dW^Q \end{aligned} \quad (1.4)$$

Proof: See Appendix A.2

Note that the dynamic dm is just a standard Brownian motion. Intuitively, m represents the distance-to-default normalized by volatility (σ). With the assumption that the barrier (B) grows at the same rate as the value of firm (V), the dynamic of distance to default is just the same random walk represented by the stochastic term of dV . In other words, with a constant expected leverage ratio, the distance-to-default is just a martingale.

Then consider dS from (1.3). Plug in $dm = dW^Q$, we get:

$$dS = \frac{1}{2} \left(\frac{\partial^2 S}{\partial m^2} \right) dt + \frac{\partial S}{\partial m} dW^Q \quad (1.5)$$

which means the dynamic of the CDS spread is governed solely by $\frac{\partial^2 S}{\partial m^2}$ and $\frac{\partial S}{\partial m}$. The values of $\frac{\partial S}{\partial m}$ and $\frac{\partial^2 S}{\partial m^2}$ can be found either by a numerical integration and differentiation of (1.2) or by analytical means (in Appendix A.3). We have seen that $S(m)$ is a one-to-one function of m . Moreover, $\frac{\partial S}{\partial m}$ and $\frac{\partial^2 S}{\partial m^2}$ are one-to-one functions of m . Thus, we can find one-to-one functions $h(S) = \frac{\partial S}{\partial m}$ and $l(S) = \frac{\partial^2 S}{\partial m^2}$, which depend only on CDS spreads themselves. We can write (1.5) as

$$dS = \frac{1}{2} l(S) dt + h(S) dW^Q \quad (1.6)$$

To infer the correlation between any two assets we first write:

$$\frac{dS_i - \frac{1}{2} l(S_i) dt}{h(S_i)} = dW_i^Q$$

where the subscript i denotes distinct asset index. The correlation between any two assets is given by:

$$\rho = \text{corr}(dW_i, dW_j) = \text{corr}(dW_i^Q, dW_j^Q) = \text{corr}\left(\frac{dS_i - \frac{1}{2} l(S_i) dt}{h(S_i)}, \frac{dS_j - \frac{1}{2} l(S_j) dt}{h(S_j)}\right) \quad (1.7)$$

Note here that correlations in the physical measure are equal to correlations in the risk-neutral measure in this simple setup. From numerical integration and differentiation and polynomial fitting, we approximate $l(S)$ and $h(S)$ as follows:

$$\begin{aligned} l(S) &= 42.6778S^5 - 18.8941S^4 + 73.9661S^3 + 1.8244S^2 + 0.5223S + 0.0013 \\ h(S) &= -4.3931S^5 + 9.6218S^4 - 7.7494S^3 - 2.9201S^2 - 0.6497S - 0.0014 \end{aligned}$$

1.3 Empirical Analysis

1.3.1 Data

We use CDS data¹ from Credit Market Analysis (CMA), acquired by CME on March 25, 2008. The data are daily ranging from January 2004 to May 2008. For this paper we use only 5-year CDS data because they are the most common and most liquid. For convenience we use only the CDS data with complete time series over the specified period. Since we want to compare asset correlations inferred from CDS spreads with equity correlations, we then consider only CDS data with matching stock returns data from CRSP. We end up with 193 firms over the period of about 4.5 years. The summary statistics of the CDS data and matching stock returns are presented in Table 3.9 and Table 3.2.

¹We thank Tom Jacobs for collecting and preprocessing the data

1.3.2 Empirical Results

From the theoretical result $\rho = \text{corr}(dW_i, dW_j) = \text{corr}(\frac{dS_i - \frac{1}{2}l(S_i)dt}{h(S_i)}, \frac{dS_j - \frac{1}{2}l(S_j)dt}{h(S_j)})$, we discretize as follows:

$$\rho = \text{corr}(\frac{S_{i,t+1} - S_{i,t} - \frac{1}{2}l(S_i)\Delta t}{h(S_{i,t})}, \frac{S_{j,t+1} - S_{j,t} - \frac{1}{2}l(S_j)\Delta t}{h(S_{j,t})})$$

We use monthly intervals for discretization (and thus $\Delta t = 1/12$) to avoid autocorrelation in CDS data. From the model we can infer pairwise asset correlations of 193 firms. We report a histogram of lower triangular entries of the correlation matrix. We then compare this result with the corresponding equity correlations from the same set of firms over the same period. The result is shown in Figure 1.4. Pairwise differences in correlations are shown in Figure 1.5.

There is a big difference between equity correlations and asset correlations as shown in Figure 1.4 and 1.5. The magnitude of the mean difference is 0.260 which is very large for correlations. If the true correlations are those implied by CDS spreads, one can be off by a large amount if one uses equity correlations in place of asset correlations to calculate default correlations. One question arises: which correlation is correct? We will address this question in the Default Simulation section.

1.4 Robustness Check

The result from the previous section depends on the correctness of the model, which, in turn, depends on the assumptions about the underlying processes of firms and debt barriers. In this section we incorporate more realistic

assumptions from recent literature about these processes. We show that the main results still hold.

Collin-Dufresne and Goldstein (2001) shows that the model for stationary leverage ratios fits the credit spread data better. Firms adjust outstanding debt levels in response to firm value, thus generating mean-reverting leverage ratios. The key difference between their model and our model is that there is no target leverage ratio in our model. In fact, for our model, if the asset volatility is constant, the log-leverage is simply a Brownian motion of the firm scaled by the standard deviation of the underlying asset. The model yields a very simple close form solution yet it can be viewed as too simplistic. In this section we start with the stationary (mean-reverting) leverage model as in Collin-Dufresne and Goldstein (2001) and derive the formula for implied correlation. Then we calibrate the model to the data and find the empirical results to compare with the previous section.

Similar to Collin-Dufresne and Goldstein (2001), we let the log-leverage l_t follow the mean-reverting process in the risk-neutral measure:

$$dl_t = \lambda(\bar{l}^Q - l_t)dt + \sigma dW_t^Q \quad (1.8)$$

where $\lambda = 0.18$, $\bar{l}^Q = -0.6556$ for investment-grade firms and -0.5556 for high-yield firms, and $\sigma = 0.2$.

The distance-to-default m is defined as $\frac{\log \frac{V}{B}}{\sigma} = \frac{-l}{\sigma}$. Thus, the dynamic of m is:

$$dm_t = \lambda(\bar{m}^Q - m_t)dt + dW_t^Q \quad (1.9)$$

Now default is defined as the first time $V_t = B_t$ or the first time $m_t = 0$. There is a closed-form solution to the first passage time density of a mean-reverting process (see Appendix A.4). Similar to the previous section, this

density is the default density and thus we can find CDS spreads as a function of the starting distance-to-default (m_0). We show the plot in Figure 1.6.

We can see that, for the same distance-to-default (m), the CDS spread from the stationary leverage model is lower than the original model without mean reversion. The result is intuitive because with mean-reversion, the high-leveraged firms will decrease the leverage going forward, thus reducing the probability of default. The model is calibrated to the parameters for investment-grade firms, which account the majority of CDS data.

The dynamic of CDS spreads is similar to the previous section, but with a slight change in the drift term. In particular,

$$\begin{aligned}
dS &= \frac{\partial S}{\partial m} dm + \frac{1}{2} \frac{\partial^2 S}{\partial m^2} (dm)^2 \\
&= \frac{\partial S}{\partial m} \left[\lambda(\bar{m}^Q - m_t) dt + dW_t^Q \right] + \frac{1}{2} \frac{\partial^2 S}{\partial m^2} dt \\
&= \left[\frac{1}{2} \frac{\partial^2 S}{\partial m^2} + \frac{\partial S}{\partial m} \lambda(\bar{m}^Q - m_t) \right] dt + \frac{\partial S}{\partial m} dW_t^Q \quad (1.10)
\end{aligned}$$

The dynamic of CDS spreads in (1.10) is similar to (1.5) except for the drift term. We calibrate the model to the data similar to the previous section but using the mean-reverting CDS spread curve instead. We find $l(S) = \frac{1}{2} \frac{\partial^2 S}{\partial m^2}$ and $h(S) = \frac{\partial S}{\partial m}$ as a function of S itself and get the following curve:

$$\begin{aligned}
l(S) &= -13273.51S^5 + 5573.63S^4 - 551.23S^3 + 53.7093S^2 - 0.0277S + 0.0019 \\
h(S) &= 391.2648S^5 - 199.1543S^4 + 33.1832S^3 - 11.8712S^2 - 0.5717S - 0.0011
\end{aligned}$$

With the dynamics of CDS spreads, the implied correlation can be cal-

culated similarly as before:

$$\begin{aligned}\rho &= \text{corr}(dW_i^Q, dW_j^Q) \\ &= \text{corr}\left(\frac{dS_i - \frac{1}{2}l(S_i)dt - h(S_i)(\lambda(\bar{m}_i^Q - m_i))}{h(S_i)}, \frac{dS_j - \frac{1}{2}l(S_j)dt - h(S_j)(\lambda(\bar{m}_j^Q - m_j))}{h(S_j)}\right)\end{aligned}$$

With the closed-form solution, we can calculate the implied correlation from CDS spreads as before. The value of \bar{m}^Q is the target distance-to-default of investment-grade firms. We can find m from the CDS spreads just by inverting the graph in Figure 1.6.

We calibrate the stationary leverage model to the same dataset. The result is similar to the original model without mean reversion. The average implied asset correlation from CDS spreads is 0.4249, in line with the result from our model without mean reversion. We also calibrate the model to the parameters for high-yield firms and the average implied asset correlation is 0.43. Thus, our main result still holds even when we adopt a more sophisticated model with stationary leverage ratios.

1.5 Default Simulation

It is hard to determine which correlation, implied correlation from CDS or equity correlation, is the correct value, since we do not know the true asset correlation in the first place. One way to determine is to check whether asset correlations from CDS spreads predict default correlations. However, it will come down to the same problem that defaults are rare and thus we will not have reliable statistical results. Another way is to match the standard deviation of yearly default rates with the true correlations. This approach

comes from the fact that correlations determine the standard deviation of default rates, but not the average. We use this approach in this section.

We take the descriptive statistics of default rates from Moody's 2011 report which has the historical data of default rates from 1920 - 2010. We focus only on B and Caa-C firms because the mean and standard deviation of default rates are large enough to simulate and compare meaningfully with the statistics of historical data. In particular, we want to match the historical statistics in Table 1.3.

We proceed as follows:

1. Find the distance-to-default (m) that matches the average default for B and Caa-C firms for the next year ($= 2N(-m)$)
2. Simulate 100 firms of the same category with the same pairwise asset correlations (ρ).
 - The underlying process of the firm follows a Brownian motion
 - Default occurs when the Brownian motion hits the barrier m
 - Use 200 intervals in 1 year and 10,000 trials
3. Find ρ that matches the standard deviation of default rates

The result is shown in Table 1.4. For B firms the implied correlation from simulation is 0.25, while for C firms it is 0.45. The result does not point to one absolute number for the true correlation. However, the implied correlation from the simulation is high, more in line with our result from CDS spreads rather than the correlation from equity returns.

1.6 Interpretation

What can explain the difference in implied asset correlations from CDS spreads and correlations from equity returns? If the equity market and the credit market are integrated and the model is correct, then the implied correlation from both markets should be the same.

One explanation can be that the bond market, or credit risk in particular, is driven by another factor apart from the systematic factors that drive equity returns. Many papers also found the same phenomena. For example, Collin-Dufresne, Goldstein and Martin (2001) found that there is a single common factor, apart from several standard economic and financial variables, that drives credit spreads. Das et al. (2007) also found that there are unobservable explanatory variables for corporate defaults that are correlated across firms. If this unobservable factor drives corporate defaults and credit spreads, then it will also drive CDS spreads. This factor can be the reason why the implied correlation from CDS spreads is higher than the correlation from equity returns. In other words, the implied correlation from CDS spreads is the result of two driving forces: the systematic factor that drives equity returns, and the unobservable factor that has been found to drive corporate defaults and credit spreads.

This explanation supports our view that the implied correlation from CDS spreads should be a better proxy for default correlations than the correlation from equity returns. Our intuition is that the credit market is closer to default and thus it will provide a more accurate estimation of default correlations. Moreover, defaults are driven by a separate unobservable factor apart from the systematic factor in the equity market. We can only detect this factor in

the credit market but not the equity market.

Another explanation is that correlation is not constant and CDS spreads reflect the correlation in the distress period. During the year 2008-2009 (during the financial crisis), the mean equity correlation is 0.412, much higher than the mean equity correlation in the previous period of 0.155. This average equity correlation is in fact closer to the mean implied asset correlation from CDS spreads ($=0.416$) during the previous period (year 2004-2008). The market values CDS contracts as if there are high correlations among their asset values. The expectation of high correlations happens just before the financial crisis. It is possible that CDS spreads reflect asset correlations during the period of market stress, which is when defaults are most likely to occur. This explanation also supports our view that the implied correlation from CDS spreads is a better proxy for default correlations.

1.7 Conclusion and Future Work

We have derived a model to infer asset correlations from CDS spreads. To the best of our knowledge, our model is the first closed-form solution that links CDS spread dynamics to asset dynamics and, correspondingly, asset correlations. The asset correlations inferred from CDS spreads are much higher than the corresponding equity correlations. The result is robust even when we use the more complicated mean-reverting leverage ratio model. Once we know asset correlations and distance-to-default, we can calculate default correlations.

We proceed to determine which correlation, implied correlation from CDS

spreads or the correlation from equity returns, is correct. Default simulation of historical data suggests that the actual correlation should be higher than the equity correlation, and more in line with our correlation from CDS spreads. The results also comply with the literature in that there is an unobservable factor, apart from the systematic factor in equity returns, that drives defaults and credit spreads among firms. It is also possible that CDS spreads reflect asset correlations during the period of market stress when defaults occur; correlations themselves can change over time, with high values during the recession period. All the evidence confirms our intuition that the implied correlation from CDS, the product closest to default, is a better proxy for default correlations than the correlation from equity returns.

To relate asset dynamics and CDS spread dynamics, our model is arguably the simplest possible. With this simple background model and a closed-form solution, it is possible to extend the model to include more sophisticated products or default barrier dynamics. Future research may also include identifying the factor that drives default correlations among firms.

1.8 Figures and Tables

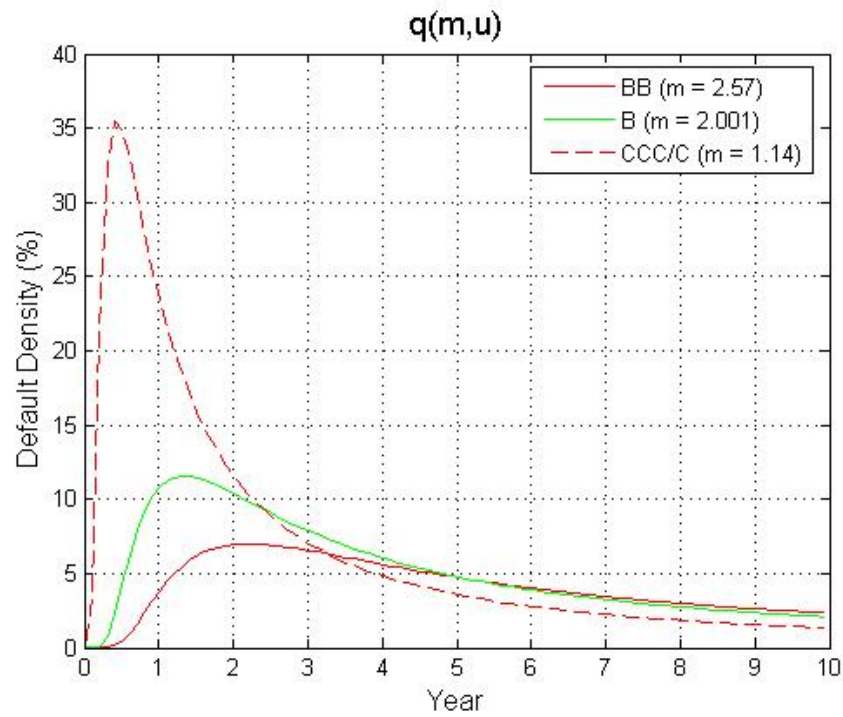


Figure 1.1: Probability density of default: $q(m,t)$. The plot shows the theoretical probability of the first passage time of asset values across debt barriers for firms with different ratings. The probability depends on the distance-to-default (m) and time (t)



Figure 1.2: Historical default density. S&P data from 1981 to 2008. This is the actual historical default rates categorized by the initial credit ratings. The plot is used to compare historical default density with the theoretical default density in Figure 1.1.

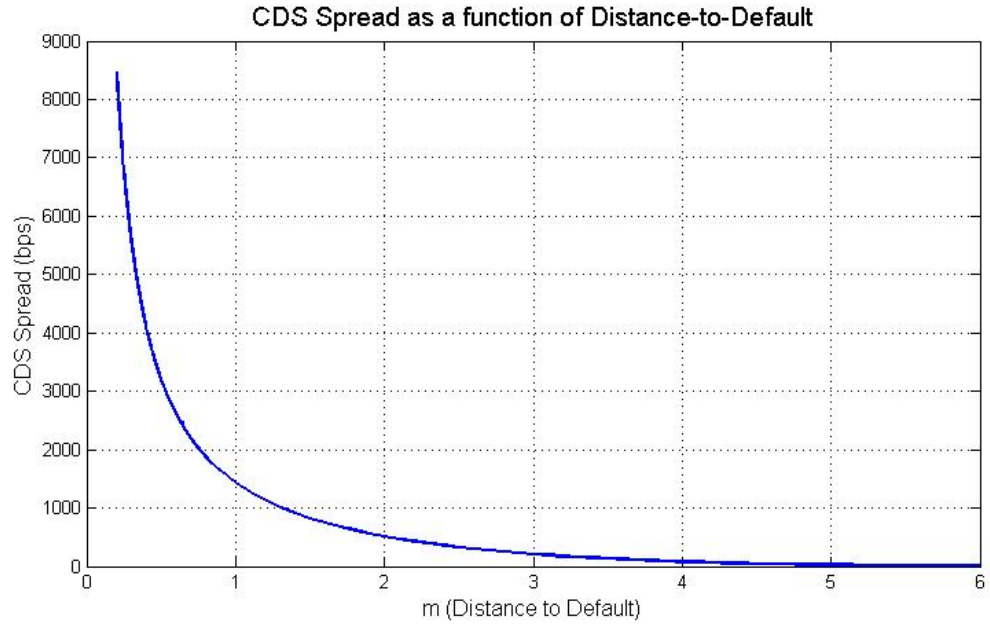


Figure 1.3: CDS spreads as a function of distance-to-default. The graph is generated from Equation 1.2 with parameters $T = 5$, risk-free rate (r) = 0.025, recovery rate (\hat{R}) = 0.4 and $a^* = 0.004$., and with the theoretical probability of default $q(m, t)$.

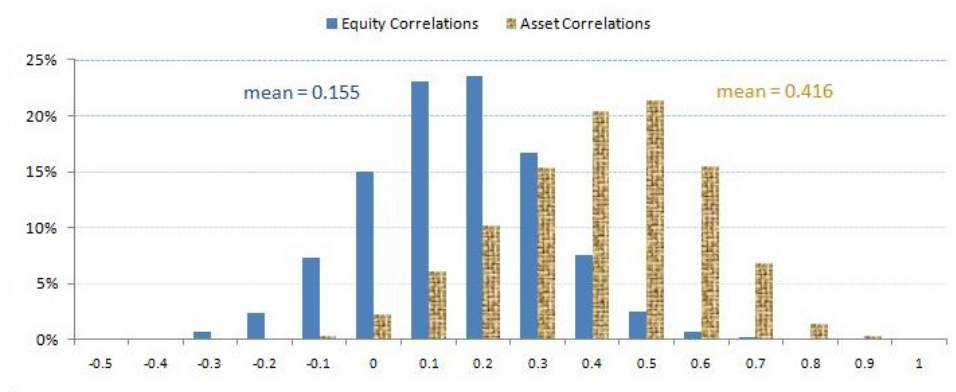


Figure 1.4: Histogram of implied asset correlations and equity correlations. The figure shows the histogram of implied pairwise asset correlations from CDS spreads using our theoretical formula (in dotted plot). The histogram of equity correlations is also shown for comparison (in solid plot). The mean asset correlation is 0.415 while the mean equity correlation is 0.155.

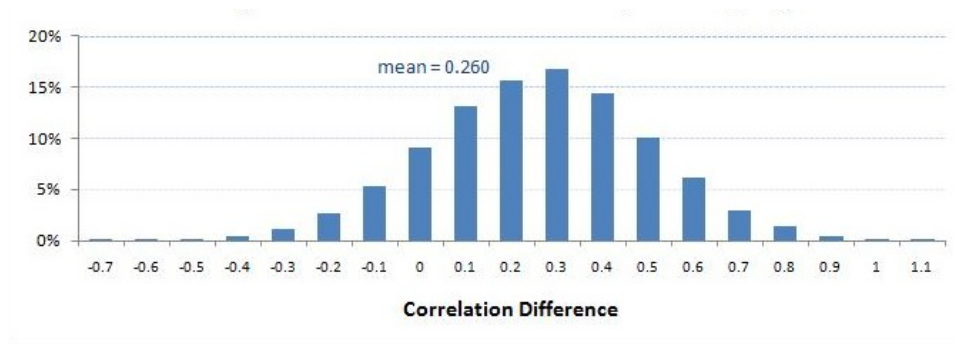


Figure 1.5: Histogram of pairwise differences in correlations (Asset - Equity). The mean difference is 0.260.

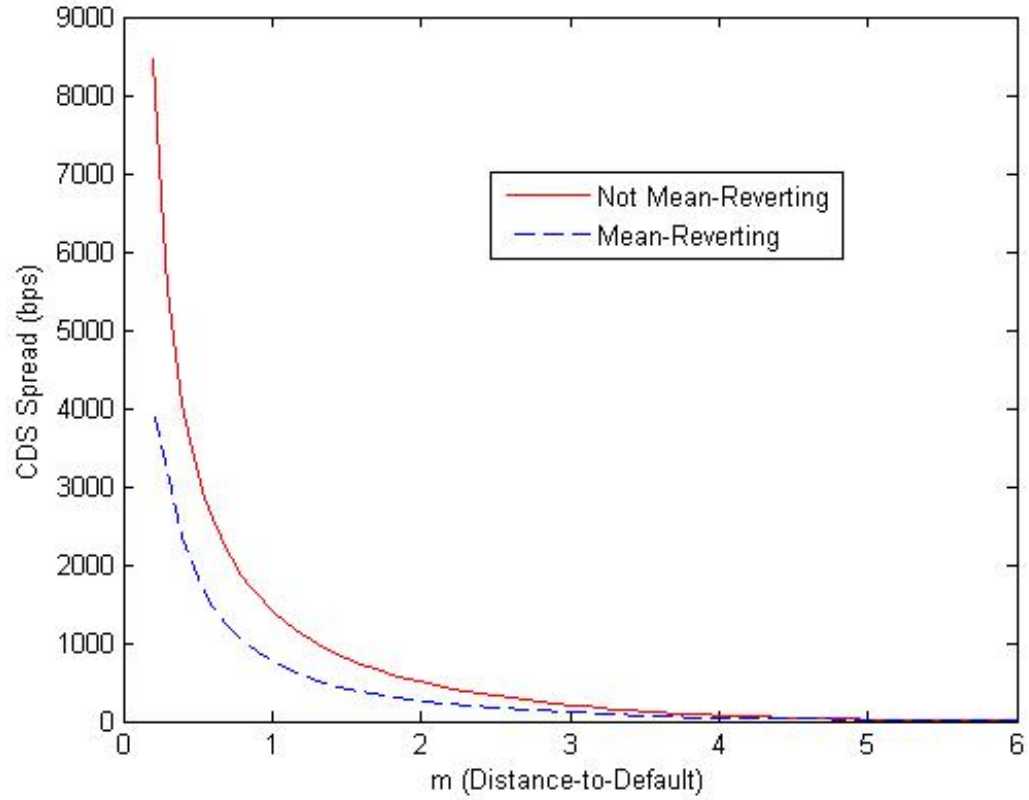


Figure 1.6: CDS spreads as a function of distance-to-default. The dashed line shows the CDS spreads corresponding to the mean-reverting leverage ratio model as in Collin-Dufresne and Goldstein (2001). The solid line shows the CDS spreads corresponding to our model with no mean reversion

Table 1.1: Summary Statistics of CDS data

The table shows the summary statistics of CDS data used in the empirical section. *CDS Spread* shows the spreads in basis points. The remaining information in the table comes from the CRSP database. *Market Cap* shows the market capitalization of firms in millions. *Stock Monthly Volatility* shows monthly volatility of the stock. *Beta*(β) shows the beta coefficient when regressing the stock returns with market returns. *SMB* and *HML* shows the regression coefficients when regressing the stock returns with the *SMB* and *HML* factors

	CDS Spread	Market Cap (M)	Stock Monthly Volatilities	β	SMB	HML
mean	79.27	27,496	0.068	0.94	0.12	0.31
median	38.60	13,524	0.062	0.85	0	0.19
min	1.5	176	0.031	-0.19	-3.05	-1.34
5% quantile	11.7	1,923	0.036	0.16	-0.87	-0.81
25% quantile	23.7	6,404	0.049	0.63	-0.33	-0.19
50% quantile	38.6	13,524	0.062	0.85	0	0.19
75% quantile	70.5	27,955	0.078	1.18	0.60	0.76
95% quantile	316.1	99,429	0.120	2.09	1.41	1.67
max	2666.5	513,362	0.229	2.77	2.44	2.79

Table 1.2: Industry Classification

This table shows the percentage of firms in the data set classified into each industry, according to the Siccodes on Kenneth French's website.

Number	Industry	Percent
1	Consumer Nondurables	7.75
2	Consumer Durables	2.87
3	Manufacturing	17.65
4	Energy	5.41
5	Hi-tech	5.69
6	Telecom	3.10
7	Wholesale, Retail	12.48
8	Healthcare	7.20
9	Utilities	8.76
10	Other: Mines, Trans,Const, Finance, etc	29.10

Table 1.3: Historical average and standard deviation of default rates

The data are from Moody's 2011 report. We focus only on B and Caa-C firms because the mean and standard deviation of default rates are large enough to simulate and compare meaningfully with the statistics of historical data.

Ratings	Average Default Rates	Standard Deviation of Default Rates
B	3.41%	4.04%
Caa-C	13.86 %	17.05%

Table 1.4: Default simulation result

We fix the average default rate of B and Caa-C firms by finding the implied distance-to-default using the theoretical formula. Then we find the implied correlation from the simulation to match the standard deviation of historical data.

Ratings	Distance-to-Default (m)	Implied Correlation
B	2.12	0.25
Caa-C	1.48	0.45

Chapter 2

The Illiquidity of CDS Market

2.1 Introduction

Credit derivatives market is relatively new. The market existed since early 1990s but grew exponentially from 2003 until the financial meltdown in 2008. Within credit derivatives, the most liquid and most traded instrument is Credit Default Swap (CDS) with 5-year maturity. Since this financial instrument is insurance on credit risk of the underlying bonds, finance academics and practitioners can extract default probabilities from its price. The credit derivatives price may even be a better indicator of default risk than the bond price itself, because the bond market is relatively illiquid and burdened with complicated contracts and maturity structures. The CDS market would be ideal for credit risk researchers, only if the market is really liquid and the spreads reflect a fair price. This paper asks this simple question: is CDS market really liquid? The answer is unfortunately no.

To test whether the CDS market is liquid or not, we need to know first

how to "measure" liquidity. In the literature, it is still not clear how to measure such quantity precisely, although we know that illiquidity will somehow affect the price. In this paper, we bypass the problem of definition and measurement of liquidity. We focus on the "event" when liquidity changes and ask the question: do CDS spreads change? Our main hypothesis is as follows. If the market condition becomes more liquid, but the CDS spreads do not change, it means that the original CDS market is already liquid, or at least as liquid as the new market. If the spreads do change, then the original CDS market is not liquid. In this paper, we found the latter to be the case.

On the Markit website, the administrator and marketer of CDX index, it says that one of the key benefits of CDX index is liquidity - wide dealer and industry support allow for significant liquidity in all market conditions. Moreover, one of the key functions of CDX index is that it enhances liquidity in the single name market - the liquidity of the index flows into the single name CDS market.

Thus, the "event" that changes liquidity here is the inclusion of CDSs into a Credit Default Swap index (CDX). The CDX index is created every 6 months and consisted of 125 CDSs. Once a CDS is included into the index, it stays there for at least 6 months until the next roll date. The CDS can also be excluded from the index if the committee decides so. The time $t = 0$ for the event is when a CDS is included into CDX and starts trading, so called the roll date. We found the cumulative abnormal changes of CDS spreads to be positive and economically high. There can be many explanations for this observation, but the reversal in the spread suggests that the most likely channel for illiquidity is that order imbalance causes price impact depending

on the direction of dealers' inventory.

For robustness check, we also conjecture that if a CDS is excluded from the index, then the liquidity will change as well and this should affect the price. We found this to be the case in the data. Moreover, the correlation between CDS and CDX, when the CDS is included, should be higher than when it is not included – the higher liquidity allows the price to adjust easily according to the information and market belief, and so the individual price co-moves more with the aggregated credit risk. We also explore this hypothesis and found positive evidence, but the evidence is not very strong. We also argue that an alternative explanation that index inclusion conveys information about the underlying credit risk is unlikely to be the case.

Recent research uses CDS spreads instead of bond yield spreads to be a proxy for default probabilities, arguing that much of bond yield spreads is due to illiquidity. Examples include Longstaff et al. (2005), Chen et al. (2007), and Huang and Huang (2003). Blanco et al. (2005) suggest that CDSs are a cleaner indicator than bond spreads, and that CDS prices are useful indicators for analysts interested in measuring credit risk. Our paper does not assume that the CDS market is liquid and will explore this issue. Several papers have explored the effect of index inclusion or adjustment on prices. Examples include Shleifer (1986) and Kaul et al. (2000). They observe excess stock returns after index inclusion or adjustment and conclude that demand curves for stocks slope down. Our paper differs in that we study derivatives products, whose net supply is zero.

The chapter proceeds as follows. Section 2.2 describes the CDX index and the inclusion and exclusion procedure. Section 2.3 describes the data for

CDS and CDX. Section 2.4 is an empirical analysis and event study. Section 2.5 is a robustness check. Section 2.6 describes the possible reasons we see an increase in CDS spreads after index inclusion. The final section concludes.

2.2 CDX Index and Inclusion and Exclusion Procedure

A Credit Default Swap index (CDX) is a credit derivative used to hedge credit risk on a basket of CDSs. The index is a standardized credit security and is more liquid and traded at a smaller bid-ask spread. There are two main families of CDS indices: CDX and iTraxx. In this paper, we only concentrate on the CDX index, especially CDX.NA.IG which contains CDSs of North American Investment Grade bonds. The index is administered by CDS Index Company and marketed by Markit Group Limited.

A new CDX index is issued every six months by Markit. The composition of the Investment Grade (IG) Index is determined based on submissions by each member that elects to participate in the determination of the IG Index and each related sub-index on a continuing basis. Each IG Index is composed of one hundred twenty five (125) entities, with equal weighting of 0.8%. Each IG Index begins on September 20 (or the next Business Day in the event that September 20 is not a Business Day) and March 20(or the next Business Day in the event that March 20 is not a Business Day) of each calendar year. Since the number of CDSs in a CDX is fixed at 125, some CDSs must be excluded before a new CDS can be included into the index. We explain below the exclusion process before the inclusion process. The information is taken

from Markit's publication, "Index Methodology for the CDX Indices, (2007)"

The polling process to decide which CDS to exclude is as follows. Ten (10) business days prior to the Roll Date of a new IG Index, the Administrator will solicit each eligible IG Member to submit a list of entities in the then current IG Index that in such Eligible IG Members judgment should not be included in the IG Index for the next six-month period based on the following criteria

1. entities for which the associated reference obligation is rated below investment grade by two of S&P, Moodys and Fitch;
2. entities for which a merger or other corporate action has occurred or been announced that renders such entity no longer suitable for inclusion;
3. entities whose outstanding debt or for which credit default swap contracts has/have become materially less liquid.

The polling process to decide which CDS to be included into the index is as follows. After CDSs have been eliminated from the index and no later than nine (9) business days prior to the Roll Date, the Administrator will determine the number of additional entities required to add to those entities remaining in the new IG Index to total one hundred twenty five (125) and will solicit each eligible IG Member to submit a list of entities. No later than seven (7) business days prior to the Roll Date (the Index Publication Date), the Administrator will publish to the public and eligible IG Members the composition of the new IG Index for that next six-month period. At such time, the Administrator will also publish to the public the current list

of eligible IG Members for the new IG Index. Two (2) business days prior to the Roll Date, the Administrator will publish to the eligible IG Members (but not the public) a draft of the annex for the IG Index and each sub-index along with the weighting and final reference obligations for each entity within the new IG Index and each new sub-index. The final annex for the IG Index and each sub-index will be published after 5:00 p.m. on the Business Day immediately preceding the Roll Date. Products based on the new IG Index will begin trading on the Roll Date.

Thus, from the Markit publication, CDSs will be decided to be excluded from the index 10 days before the new roll date. The CDSs that will be included into the index will be published to the public no later than 7 days before the new roll date. These dates are important when we do the event study.

A new development has been made to the index procedure. According to the Markit's "CDX and LCDX Rules 2012", the determinant of inclusion and exclusion is the liquidity of the individual names, as published on the DTCC Trade Information Warehouse (market risk activity for the Top 1000 names). In particular, the most liquid names will be included into the index and the most illiquid will be dropped. However, this guideline on liquidity did not exist in the prior CDX and LCDX rules in 2007 when our data concern.

2.3 Data

2.3.1 CDS Data

We use CDS data from Credit Market Analysis (CMA), acquired by CME on March 25, 2008. The data are daily ranging from January 2004 to May 2008. We use only 5-year CDS data because they are the most common and most liquid. One potential flaw of using daily data is that CDS spreads can have high autocorrelations. To get a clean result, we use monthly data in the regression analysis. However, for the event study section, we need daily data to draw conclusion, while monthly data are too crude to yield any meaningful result.

There are 24 firms that are included into the index during our study period. The summary statistics of the CDSs of these firms are shown in Table 3.9. The mean and median of the CDS spreads are within a reasonable range. However, the standard deviation of the spreads is high compared to the average. This can be due to the fact that our data end just before the credit crisis. As the economy approached the crisis, the CDS spreads became highly volatile.

It may be of interest to see how our sample CDSs are classified into different industries, although we do not have any link between CDS spreads and the industry in this paper. The industry classification is shown in Table 3.2.

The data are biased towards the wholesale/retail industry. Many firms are also classified as "Other" (Industry 10), which do not have shared characteristics. We wish to have a more balanced dataset with equally weighted

firms in each industry. On the other hand, in this paper we are not concerned with any link between industry characteristics and CDS spreads. The data should be fine for our purposes.

2.3.2 Inclusion Dates

In our study period, there are 8 inclusion dates as follows:

1. Inclusion Date 1 = 23Mar2004
2. Inclusion Date 2 = 21Sep2004
3. Inclusion Date 3 = 21Sep2005
4. Inclusion Date 4 = 21Mar2006
5. Inclusion Date 5 = 21Sep2006
6. Inclusion Date 6 = 21Mar2007
7. Inclusion Date 7 = 21Sep2007
8. Inclusion Date 8 = 21Mar2008

The period spans from the beginning of 2004 to the middle of 2008, which is the period of our CDS data. Most inclusion dates are 6-month apart, except for Inclusion Date 2(ID2) and Inclusion Date 3(ID3). Between ID2 and ID3, there is no CDS in our sample that is included in March of 2005. Note also that ID6, ID7, ID8 fall in the period when the economy was approaching the financial crisis. This explains the high volatility of CDS spreads in our sample.

2.4 Regression Analysis and Event Study

2.4.1 Regression Analysis

The first test to run is to test whether there is an average change in the spreads at all when the CDS is included into the index. A simple test is to use a dummy variable $\mathbb{1}_{[Included]}$, which is equal to 1 when a CDS is included into the index and 0 otherwise. We first report this simple regression result in Table 3.4.

This regression is still preliminary. It does not include any other factors that may affect CDS spreads. The R^2 is very low. Yet we see that the inclusion dummy is statistically significant and economically large. From this regression, if a CDS is included into the CDX index, the average spreads go up by 60 basis points. The coefficient is rather too high and calls for more control variables.

A number of other factors determine CDS spreads and should be included in the regression. We follow closely the paper by Cossin and Hricko (2001). They identify the following determinants of CDS spreads:

1. Credit Ratings: These are the most widely used measure of credit risk. We use S&P ratings available from the Compustat database. The credit ratings range from AAA to C. We assign a linear scale for each rating, i.e., AAA = 1, AA+ = 2, ..., C = 17. It may seem doubtful that ratings will affect CDS spreads in this linear fashion. However, we simply follow the same methodology as Cossin and Hricko (2001) and this method is also prevalent in the literature. Moreover, it is not clear whether (or which) a more sophisticated function of ratings would

produce a better fit. Furthermore, credit ratings are quite sticky, i.e., they do not change very often. Thus, in our regression which focuses on the event that the CDS is included into an index, credit ratings do not appear to have a lot of effect on the coefficient of interest.

2. Interest Rates: This factor affects the discount factor and the probability of default in the risk-neutral measure. We use both short-term (3-month) and long-term (5-year) treasury bills rates in the analysis.
3. Leverage: This factor indicates how close the firm's value is to the debt value. This factor is directly related to the probability of default. Another important factor is the asset volatility which is unobservable. If we assume that the asset volatility is constant for each firm, then the volatility effect can be controlled by the firm's fixed effect. Thus, we will use only leverage for each firm, and then use firm's fixed effect to control for other factors including asset volatility. Leverage is calculated by the ratio of Total Debt to Market Asset, where

$$\begin{aligned}\text{Total Debt} &= \text{Long Term Debt (DLTTQ)} + \\ &\quad \text{Debt in Current Liabilities (DLCQ)} \\ \text{Market Asset} &= \text{Asset (ATQ)} + \text{Market Equity (CSHOQ*PRCCQ)} \\ &\quad - \text{Shareholder's Equity (SEQQ)}\end{aligned}$$

The variables in parentheses indicate the corresponding variable names from the Compustat database. This calculation of leverage is standard.

4. Index Returns: This factor indicates the state of the economy, which should affect the wellness of the business, and thus the probability of

default. We use S&P500 index monthly returns available from the CRSP database.

There are 5 columns in Table 2.4, with different regression covariates. After controlling for other factors, the inclusion dummy variable is still statistically significant. Moreover, the economic magnitude is high. If we look at the first 3 columns, where the dependent variable is the CDS spreads, the coefficient for the inclusion dummy is about 30 basis points. This means that if a CDS is included into an index, then the spread increases on average by 30 basis points. This magnitude is high, although it is less than 60 basis points in the first regression. Now we discuss each column in Table 2.4 in detail.

The first column is a linear regression between CDS spreads and all control variables, all in linear scale. We use firm's fixed effect in the regression. All factors are highly significant, and the sign is in the direction that we expect. The inclusion dummy is significant.

The second column is similar to the first column, but we change the scale of leverage to a log scale. In this paper we do not assume a particular form of asset model, and thus we show that any specific functional form does not matter to our main result. In this case, if we let CDS spreads depend on the log of leverage instead of leverage itself, as in the second column, the $\log(\text{leverage})$ variable is highly significant. The inclusion dummy is still significant. However, the R^2 appears to be a little lower than the first column.

In the third column we turn back to a linear leverage model. But we change the interest rates to be in a log scale. The result is that the log of interest rates, both short and long term, are highly significant. The R^2

is higher than the first column with linear interest rates. The inclusion dummy is still significant, although the magnitude is slightly less than the first column. From the first three columns, we may conclude that linear leverage and log interest rates fit best with the CDS data. The inclusion effect is significant throughout.

In the fourth column we change CDS spreads to a log scale. The covariates are the same as in the first column. All factors are significant in this regression, but R^2 increases significantly from the first three columns. The inclusion dummy remains significant.

The fifth column is similar to the fourth column, but we change leverage into a log scale. All factors are highly significant, but the R^2 appears to be slightly lower than the fourth column. The inclusion dummy is significant.

Thus, overall, the inclusion dummy variable is significant, regardless of the functional form of the covariates. From the table, we can also conclude that the linear covariates fit best to the log of CDS spreads.

2.4.2 Event Study

From the regression analysis, we see that inclusion does affect the CDS spreads. In this section we test our hypothesis with another approach. The main hypothesis is that the liquidity of CDS changes when the CDS is included into an index, and this in turn affects the CDS spreads. Since the liquidity changes when an event occurs, another approach is to do an event study. We let $t = 0$ when a CDS is included into the CDX index. We then look at the Cumulative Abnormal Changes (CAC) around the event when $t = 0$ for each included CDS. The CAC is simply the sum of Abnormal

Changes (AC) within the same rating up to time t . Suppose there are N firms with the same rating as firm i . Starting from $t = -k$, The CAC for firm i at time t is defined as:

$$AC_{i,t} = (CDS_{i,t} - CDS_{i,t-1}) - \frac{1}{N - n} \sum_{j=1, j \text{ not included}}^N (CDS_{j,t} - CDS_{j,t-1}),$$

j and i have the same rating,

n is the number of included firms with the same rating

$$CAC_{i,t} = \sum_{s=-k}^t AC_{i,s}$$

Finally we find the Average of CAC (ACAC) of all companies that are included in the index. Suppose there are M firms (in this case $M = 24$) that are included into the index, then

$$ACAC_t = \frac{1}{M} \sum_{i=1}^M CAC_{i,t} \quad (2.1)$$

Since the index rolls every 6 months, we do an event study from $t = -125$ to $t = 125$. This corresponds to 6 months before and after the inclusion date ($t = 0$). We count only trading days without weekends. The ACAC is shown in Figure 2.1

From Figure 2.1, we can see that the ACAC is increasing around time $t = 0$. In particular, if we look at the period around 1 month before and after the inclusion ($t = -20$ to $t = 20$), we see a clear trend of positive ACAC. Note that the CDSs are announced to be included into the index at $t = -7$. The information about inclusion is incorporated into the price before the inclusion date. This graph shows that there is an abnormal change in the spreads when the CDS is included into the index, in a positive direction.

Thus, we can conclude that the CDS market is not liquid from this event study. Note that from the hypothesis, we do not specify that the change should be positive or negative. We will discuss about possible reasons for the observed positive spread changes in the robustness check section.

The tail of the graph is very volatile. This is because there are many CDSs in our sample that are included just before the credit crisis. There are 12 CDSs that are included into the CDX after March, 2007. The credit crisis drives the CDS spreads high and volatile. However, if we look at a short period around the inclusion date ($t = -20$ to $t = 20$), we can see the inclusion effect without the effect from the credit crisis.

We compute the average of ACAC before and after inclusion. The average of ACAC before inclusion is 9.86, while the average of ACAC after inclusion is 36.46. The difference between the two is $36.46 - 9.86 = 26.6$. This is in line with our regression analysis before, where the coefficient of inclusion dummy is about 30. Thus, the event study result does comply with our regression result. However, we can see that the regression coefficient is in fact too high; the highly volatile period at the tail is due to the credit crisis, and not due to the inclusion effect. The inclusion effect is still visible around the inclusion date, but the effect may not be as high as 30 basis points as our regression analysis suggested.

The ACAC may not be an absolute measure of abnormal spread changes. The average may be biased towards one outlier that outweighs other CDSs. We consider another measure of abnormal spread changes: Median Cumulative Abnormal Changes (MCAC). Instead of finding the average of CAC as in Eq (2.1), we find the median of CAC. The result is shown in Figure 2.2.

The dashed line is MCAC, while the solid line is ACAC. The MCAC is much less volatile and much lower than the ACAC counterpart. This suggests that indeed the ACAC is driven by a few highly volatile CDSs. However, looking at the MCAC around the inclusion, we still see a positive inclusion effect. In particular, from $t = -8$ to $t = 28$, or around 1-2 weeks before the inclusion up to 1 month after inclusion, the median abnormal spread changes are positive. The magnitude is rather small, but nonzero. The time period is also in line with the CDX inclusion procedure, where a CDS is announced to be included into the index 7 days before the roll date ($t = -7$). Thus, even after using the median as a measure of abnormal changes, we can still conclude that there is an abnormal spread change when a CDS is included into the index.

The window we consider in the event study is quite long and many things can happen during the 6-month period before and after inclusion. We now consider a shorter window and calculate ACAC and MCMC using the same procedure, only with $t = -50$ to $t = 50$. The result is reported in Figure 2.3.

Using the 50-day window, we can see the abnormal CDS spread changes more clearly. The spread starts to increase 10 days before the inclusion date. The change can go up to 5 basis points for the median and 15 basis points for the mean, and then reverts back slowly to zero. This reversal helps distinguish many explanations for the spread change, which we will discuss more in the Interpretation section.

2.5 Robustness Check

In the previous section we conclude that the CDS market is not liquid because we observe abnormal spread changes when a CDS is included into the index. In this section we discuss some robustness tests that we have not addressed in the previous section. We will see that all other tests do support our evidence in the previous section.

2.5.1 Exclusion Effect

If there is an inclusion effect, then there should be an exclusion effect too. Once a CDS is included into the index, it can be excluded in the next roll dates if the committee decides so. In our sample, there are 6 CDSs that are dropped out of the index. We do the same event study but with exclusion as an event. The time $t = 0$ indicates the date when the CDS is dropped out of the index. We define ACAC in a similar manner. The result is shown in Figure 2.4.

We can see the same pattern here, but in a reverse direction. Around the date of exclusion, $t = -6$ to $t = 10$, the ACAC declines (almost) monotonically. This time period overlaps with the time when the committee decides which CDS to be excluded from the index ($t = -10$). The abnormal changes increase significantly long before the exclusion, and then decline steadily through the exclusion date until long after the exclusion. Specifically in this graph, the abnormal changes increase and peak at $t = -43$ and then decline until date $t = 70$. The original increase of abnormal changes may be the reason why the committee decides to drop these CDSs out of the index in the

first place. Although the graph does not show a decline sharply only around the exclusion date, it does not contradict our conjecture that inclusion effect increases the spreads, while exclusion effect can decrease the spreads.

2.5.2 Correlation Change

Another way to detect the change in CDS liquidity is through correlations. In the previous sections, we see the spread changes through the "level". However, liquidity can also affect the correlation of CDS spreads with the aggregate credit condition. For example, if the CDS market is already liquid, then investors can buy and sell credit protection through the CDS market and CDS spreads will quickly reflect the credit condition. Inclusion of CDS into the CDX index should not affect the correlation of CDS and the credit condition, which can be measured by the CDX index. However, if the CDS market is not liquid, but the CDX index is more liquid, when a CDS is included into the CDX, then the CDS spreads will have to adjust through index arbitrage. The correlation between CDS spreads and the CDX index after inclusion will be higher than before because CDS spreads can adjust more quickly to the credit condition. In summary, our hypothesis is that if the CDS market is illiquid, then when a CDS is included into the CDX index, the correlation between CDS and CDX will increase. The other direction should also be true. First we test the hypothesis for index inclusion and then move on to index exclusion.

Index Inclusion

We test the hypothesis by finding the correlation between the CDS spreads and the most recent CDX index. In order to be comparable, we use the same most recent CDX index before and after inclusion. For example, if a CDS for firm A is included into CDX.NA.IG5, which has just been created, then the correlation before inclusion is $\text{corr}(\Delta\text{CDS of A}, \Delta\text{CDX.NA.IG4})$ and the correlation after inclusion is also $\text{corr}(\Delta\text{CDS of A}, \Delta\text{CDX.NA.IG4})$. The latter correlation should have been $\text{corr}(\Delta\text{CDS of A}, \Delta\text{CDX.NA.IG5})$, but in that case we won't be able to compare the result with the correlation before inclusion, when CDX.NA.IG5 does not exist yet. The correlations are calculated over the period of 6 months before the inclusion and 6 months after, corresponding to the rolling period.

We use weekly data for this test. We use CDX data from Bloomberg. The CDX data are only from CDX.NA.IG5 to CDX.NA.IG9. The data before the index 5 are too sparse to yield meaningful results, while the data after the index 9 are over the time period for our CDS data. Out of 24 CDSs, only 10 have enough data to yield meaningful results when calculating the correlation. We report the result in Table 2.5.

Out of 10 CDSs, only 5 have increased correlations after being included into the index. This result seems inconclusive. However, looking at the standard deviation of ΔCDS before and after inclusion, we see some patterns here. Some CDSs have much higher volatility after being included into the index. This increase in volatility can be explained in two ways. First, after being included into the index, the CDS becomes more liquid and the volatility reflects the volatility of the market. In other words, it is easier to buy and

sell credit protection and so the price adjusts according to the market. When the market is not liquid, then the price doesn't reflect the fundamentals and can be sticky. Second, the increase in volatility has nothing to do with the inclusion effect. It is due to the credit crisis and macro economy in general. The first and second explanations are both valid and we do not have a way to separate them. We then consider each case and its implication to our hypothesis that the CDS market is illiquid.

1. If the increase in volatility reflects an increase in CDS liquidity, then we should count the increase in volatility itself as an indicator of the change in liquidity. We may not expect the correlation to increase in this case, because higher volatility in the denominator will naturally decrease the correlation coefficient. In Table 2.5, in the last column, we consider the CDS as high volatility after inclusion if the standard deviation of ΔCDS after inclusion is more than 3 times higher than before. We then count these high-volatility CDSs as reflecting the change in liquidity. Thus, overall, there are 5 CDSs that change liquidity by high volatility after inclusion, and 3 more CDSs that have higher correlations after inclusion. A total of 8 out of 10 CDSs shows an increase in liquidity after inclusion.
2. If the increase in volatility reflects the credit crisis or other factors unrelated to liquidity, we should take these CDSs out of our consideration when considering the correlation change. Again, we cannot compare the correlation meaningfully if the standard deviation before and after are not comparable. We exclude the 5 CDSs with high volatility from our consideration. Out of the remaining 5 CDSs, three have higher

correlations after the inclusion. In this case, the evidence is not very strong.

In sum we see that the evidence of correlation change after inclusion is not very strong, though not in contradiction with our hypothesis.

Index Exclusion

In this section we test the hypothesis the other way around for index exclusion. We follow the same methodology as for the index inclusion. There 6 CDSs that are excluded from the index that have enough data to yield meaningful results. The result is reported in Table 2.6.

The change in correlations is also inconclusive for the exclusion case. Out of 6 CDSs that are excluded, three have lower correlations and three have higher correlations. Note that the standard deviations of ΔCDS do not change significantly before and after exclusion. The last column of Table 2.6 shows the ratio of the standard deviations of ΔCDS after exclusion to those before exclusion. Most numbers lie between 0.5 and 1.5 and we can conclude that there is roughly no change in volatilities of CDSs after exclusion. This observation is in contrast with the inclusion case. In Table 2.5, about half of the CDSs that are included into the index have higher volatilities. The exclusion and inclusion dates are in the same period, and so the difference should not come from macro economic factors. Nevertheless, for the exclusion effect, the liquidity change is also inconclusive.

2.5.3 Bid-Ask Spreads As A Measure of Liquidity?

The first test of liquidity should have been bid-ask spreads. However, from the data we do not see a significant change in the bid-ask spreads of CDS before and after inclusion. Before the event, the average bid-ask spread is 6.2 basis points, and after the event, 6 basis points. Thus, looking at bid-ask spreads will also provide inconclusive evidence of the liquidity of the CDS market.

2.5.4 Alternative Explanations: Information or Demand Curves Sloping Down?

Another argument for the spread changes may be that index inclusion conveys information about the underlying CDSs, and triggers the change of market perception of the underlying credit risk. We argue that the information story is unlikely. If the inclusion conveys information and we observe an increase in spreads, it means that the index committee specifically selects high-risk CDSs to be included into the index. This is not likely to be the purpose of the index committee. The purpose of the index is to track the performance of various segments of credit derivatives and to provide a benchmark for funds that invest in similar products.

An alternative explanation is that index inclusion increases the demand for individual CDSs. As demand increases, prices go up and this event shows that demand curves slope down. The argument is similar to Shleifer (1986). The paper investigates the positive abnormal returns to the stock index inclusion and concludes that, because of the higher demand from index funds, the

demand curves slope down. Our paper is different in that our interested products are financial derivatives. Although there may be higher demand from index inclusion, derivatives products are priced by a no-arbitrage principle rather than the supply and demand effect. In other words, the derivatives products are in zero net supply. In a perfectly liquid market, derivatives are priced to match the expectation of the payoff from the underlying financial products. If the CDS market is liquid, then we should see the CDS spreads unchanged after inclusion.

2.6 Interpretation

In the hypothesis we only look for abnormal changes of the spreads when liquidity changes. We do not specify whether the abnormal changes should be positive or negative. It turns out that the abnormal changes are positive when a CDS is included into the index. We offer here possible explanations.

A CDS is a contract between a buyer and a seller and the price, in a liquid market, should reflect the market perception of credit risk. As a derivative product, a CDS should be priced by a no-arbitrage principle reflecting the credit risk of the fundamentals. When the market is not liquid, the price may not perfectly reflect the credit risk of the underlying asset, but may also depend on the supply and demand from the buyer and the seller or the microstructure of the market.

In our case, we find an increase in the CDS spreads after being included into the CDX index. The first explanation is that the demand for credit risk from the buyer is higher than the supply from the seller. In an illiquid

market, the buyer is willing to pay a higher price but there are not enough sellers to sell credit protection. Or the market is illiquid such that it is costly for the buyer who is willing to pay for a higher price, the marginal buyer, to meet with the seller, and so the transactions do not occur. For example, there may be a retail creditor who wants to buy credit protection but is not willing to engage in the OTC negotiation process with large investment banks. In sum, there are not enough players in the market that would drive the price to the equilibrium. The CDS price in the illiquid market thus appears lower than the credit risk of the fundamentals.

When a CDS is included into the CDX index, the liquidity changes. Wide dealer and industry support allow for significant liquidity in the market. Buying credit protection becomes easy in the standardized market. Now the buyer can buy credit protection through the index with standardized contracts. The seller is also more willing to sell credit protection through the CDX market. The high demand from the buyer is met by the supply and there are enough players in the market to drive the price to the equilibrium. The price of individual CDSs need to adjust to the fundamentals, otherwise there will be an index arbitrage opportunity. The CDS spreads increase accordingly.

The second explanation is the order imbalance of end users causes price impact, which depends on the direction of the dealers' inventory. This argument is in line with Shachar (2012), Madhavan and Schmidt (1993) and Hasbrouck and Sofianos (1993). The basic mechanism is as follows. Customers buy credit protection on the CDX index from the seller. The seller hedges by buying protection on CDX components, i.e., individual CDSs.

Thus, there is high demand for individual CDSs in the index. The dealers who sell CDSs have to manage their inventory and this creates price impact on the CDS spreads. In particular, with high demand, the dealers increase the quoted prices so as to deter additional buyers and bring their inventory to the preferred position. A closer look at the market microstructure level should provide a clearer explanation for this argument.

The third explanation is that the credit risk actually increases with more liquidity of the market. The argument is along the same line with Bolton and Oehmke (2011). If the single name CDS market is illiquid, then it is hard for creditors to seek default protection. It may not be of the best interest of creditors to let the firms default. With the CDX index, creditors can easily protect themselves from credit risk. Thus, they no longer have an incentive to negotiate the firms out of default in distress periods. Credit risk increases accordingly.

From our empirical evidence, it seems the second explanation about order imbalance and dealers' inventory is the most likely. The increase in CDS spreads is impermanent and this should rule out the last explanation. Moreover, the correlation change is inconclusive and bid-ask spreads do not change and this evidence should rule out the first explanation.

2.7 Conclusion and Future Work

We found evidence that the CDS market is not liquid. We focus our study on the event when liquidity changes and observe the effect on the CDS spreads. Our main hypothesis is that if the market condition becomes more liquid,

but the CDS spreads do not change, it means that the original CDS market is already liquid. If the spreads do change, then the original CDS market is not liquid. The event is when a CDS is included into the CDX index, which is more liquid than the original CDS market. We found that the CDS spreads increase when the CDS is included into the index. The empirical results from regression analysis and event study both point to the same conclusion. Since the spreads change when the event occurs, we can conclude that the original CDS market is not liquid.

We also use correlations as a measure of change for CDS spreads. Our hypothesis is that the correlation between CDS spreads and the CDX index should increase when the CDS market is more liquid because it is easier for the spreads to adjust for different credit conditions. However, the empirical result on correlation changes is still inconclusive. Moreover, bid-ask spreads of CDS prices do not change after the inclusion.

Our hypothesis only indicates that there should be a change when liquidity changes but does not specify the direction. We found that the change is positive when a CDS is included into the index. One explanation is that demand is higher than supply in the newly created market. When the market is illiquid, there are not enough players in the market to drive the price to the equilibrium, and so the price does not reflect the underlying credit risk. When liquidity increases, the high demand from retail marginal buyers is met with the supply, and so the price is higher than before. Another explanation is that order imbalance of end users causes price impact because of the dealers' inventory. High demand for credit protection forces dealers to set high prices for CDSs to deter future buyers and bring their inventory to the

preferred position. The high demand is created when a CDS is included into the CDX index which makes it become more popular. The final explanation is that credit risk actually becomes higher after inclusion. Before inclusion, it is hard for creditors to buy credit protection and thus they are more willing to negotiate with distressed firms. After inclusion, creditors gain easy protection and have more incentive to let the troubled firms default. The impermanent price impact and inconclusive results on correlation changes and bid-ask spreads suggest that the most likely explanation is the second one on order imbalance and dealers' inventory.

It is unfortunate that the CDS market is illiquid, and so the spreads many not reflect "pure" credit risk. The future work is to identify and disentangle the illiquidity component from the credit risk component in CDS spreads. The survey paper by Brigo, Predescu and Capponi (2010) may be a good place to start. However, the definition and precise measurement for liquidity remains elusive in the finance literature.

2.8 Figures and Tables

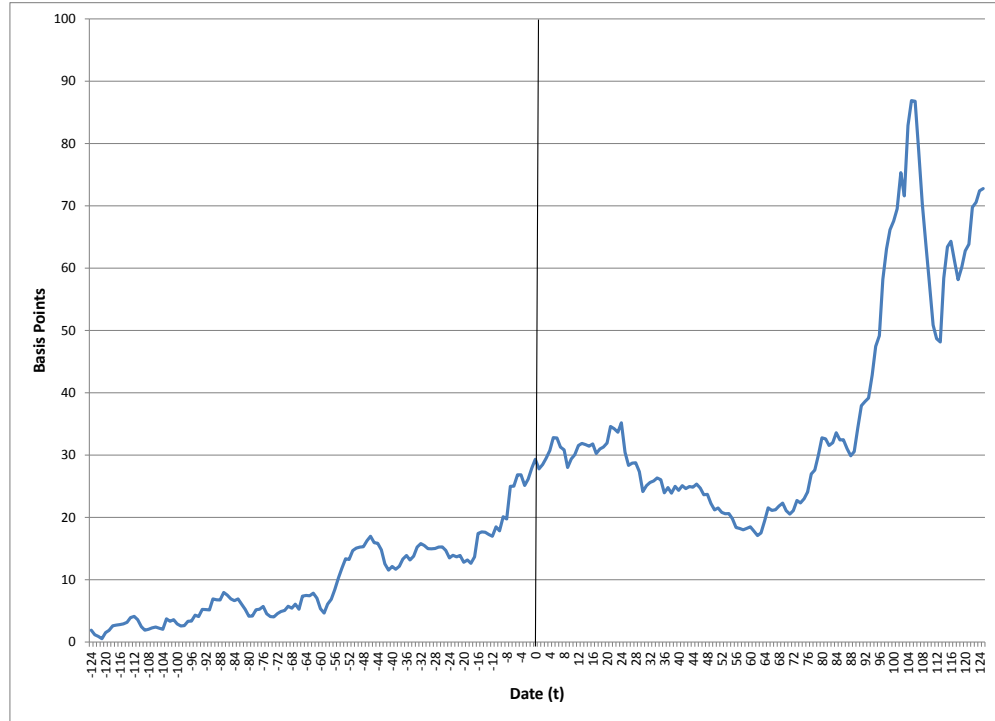


Figure 2.1: Average Cumulative Abnormal Changes of CDS spreads around the inclusion date. The cumulative abnormal changes are calculated from the difference between the changes of CDS spreads of the included firm and of other firms in the same rating. We then find the average of the cumulative abnormal changes of all 24 included firms in our sample.

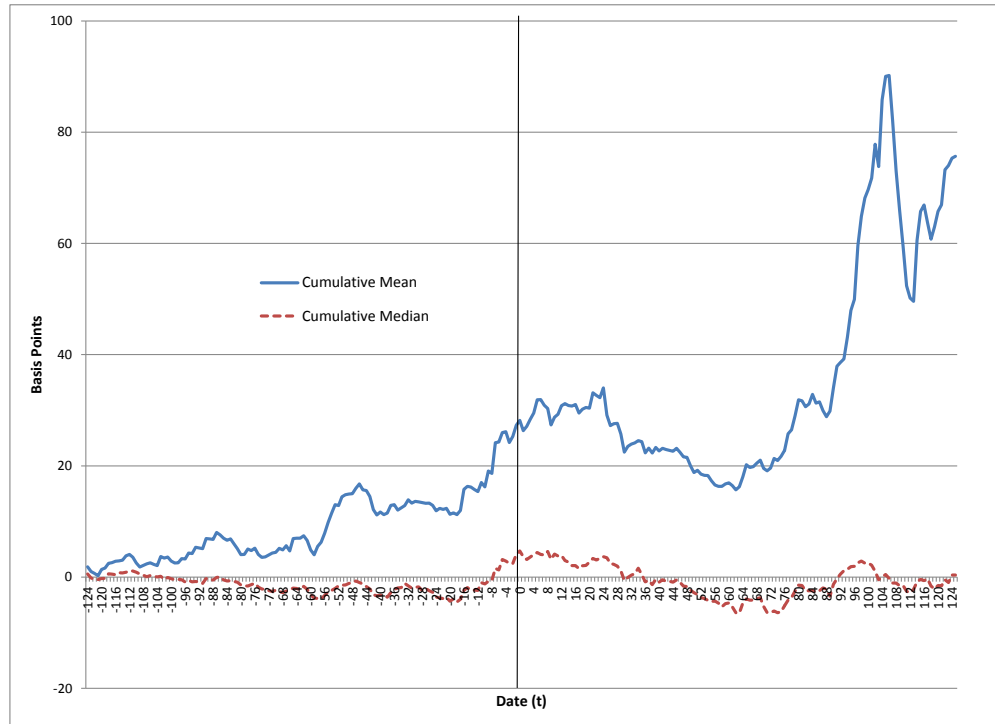


Figure 2.2: Median Cumulative Abnormal Changes and Average Cumulative Abnormal Changes. The cumulative abnormal changes are calculated from the difference between the changes of CDS spreads of the included firm and of other firms in the same rating. We then find the median and the average of the cumulative abnormal changes of all 24 included firms in our sample, shown in dashed and solid line respectively.

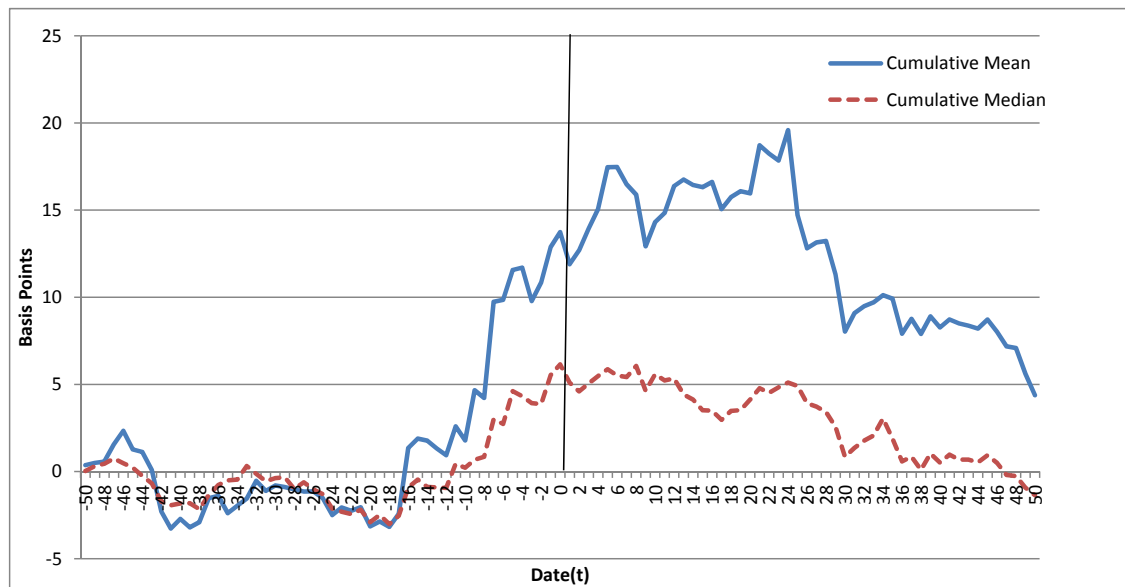


Figure 2.3: Median Cumulative Abnormal Changes and Average Cumulative Abnormal Changes, using a 50-day window. The cumulative abnormal changes are calculated from the difference between the changes of CDS spreads of the included firm and of other firms in the same rating. We then find the median and the average of the cumulative abnormal changes of all 24 included firms in our sample, shown in dashed and solid line respectively.

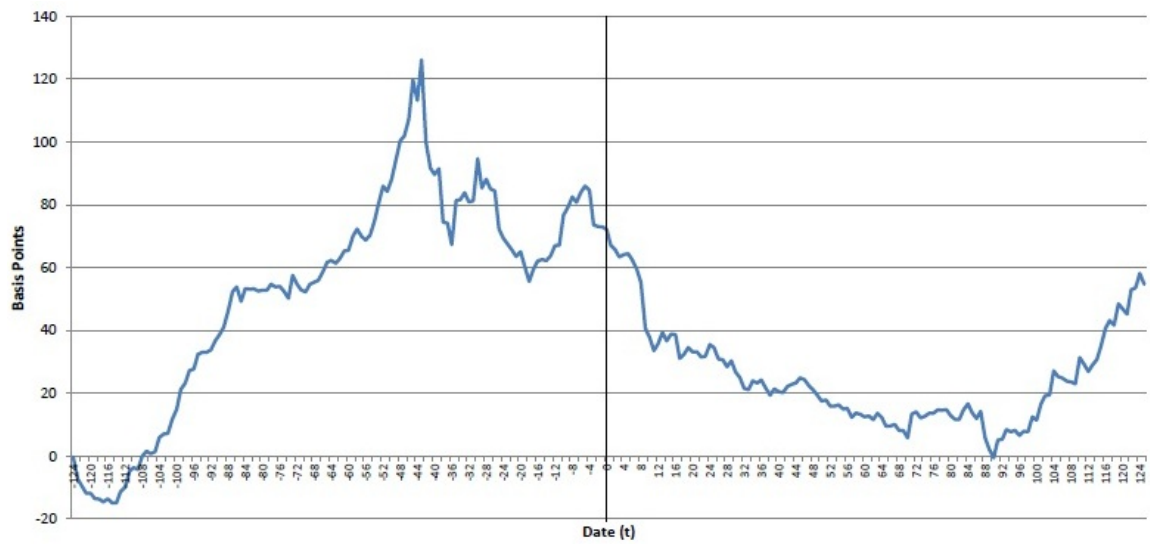


Figure 2.4: Average Cumulative Abnormal Changes for excluded CDSs.

There are 6 CDSs in our sample that are dropped out of the index

Table 2.1: Summary statistics of CDS spreads

There are 24 CDSs that are included into the index during our study period, January 2004 to May 2008. The statistics are calculated from monthly data.

Trading Symbol	N	Mean	Median	Min	Max	Standard Deviation
AZO	52	67.48	68.95	27.7	126.8	24.68
BC	52	68.43	54.65	22.9	280	57.17
BSX	52	75.12	42.80	16.2	310	74.51
CAH	52	46.03	35.35	16.5	135.8	26.24
CTL	52	75.66	71.10	39.5	148.3	24.37
DRI	52	54.22	43.70	26	201.1	38.00
FO	52	50.49	45.75	10.5	206.7	40.60
GCI	52	52.17	38.80	19.8	245	46.08
GPS	52	91.57	92.50	45	142.2	22.39
HD	52	31.34	12.65	8	196.4	43.20
HOT	52	128.95	120.25	62.8	267.5	43.50
JCP	52	94.86	81.45	35.5	238.8	51.83
KSS	52	44.20	31.90	19	163.8	36.54
LEN	52	155.63	73.15	44.5	844.8	198.18
LIZ	52	79.90	50.15	30	385	85.58
MAS	52	66.12	47.20	27.5	293.2	57.56
MCK	52	44.23	36.30	17	99	20.68
MDC	52	103.24	89.50	49.7	210	39.26
MMC	52	58.11	49.20	29	152.3	26.01
OLN	52	89.79	75.60	47.5	223.1	37.74
RDN	52	173.41	52.70	29.7	1040.1	282.53
RSH	52	113.69	119.75	27.5	319.6	64.75
TIN	52	88.83	77.60	34	288.3	52.54
WEN	52	112.11	106.25	30.5	337.5	73.29
Total	1248	81.90	60.00	8	1040.10	90.51

Table 2.2: Industry classification

The included firms are classified into each industry according to the Siccodes on Kenneth French's website.

Number	Industry	Number of firms
1	Consumer Nondurables	2
2	Consumer Durables	0
3	Manufacturing	4
4	Energy	0
5	Hi-tech	0
6	Telecom	1
7	Wholesale, Retail	10
8	Healthcare	1
9	Utilities	0
10	Other: Mines, Trans, Const, Finance, etc	6

Table 2.3: Regression on inclusion dummy variable

The table shows the OLS regression result of CDS spreads on the inclusion dummy variable (=1 when the CDS is included into the index). Standard errors are in parentheses.(***significant at 1% level, ** significant at 5% level, * significant at 10% level)

Explanatory Variables	(1)
Constant	56.34*** (3.18)
Inclusion	60.41*** (4.90)
Obs	1248
R^2	0.11

Table 2.4: Regression with control variables

The table shows OLS regressions with many well-known financial variables that affect CDS spreads. In the first 3 columns, the dependent variable is CDS spreads. In the last 2 column, the dependent variable is $\log(\text{CDS spreads})$. Firm's fixed effect is applied to all regressions. Standard errors are in parentheses. (***)significant at 1% level, ** significant at 5% level, * significant at 10% level)

Explanatory Variables	(1)	(2)	(3)	(4)	(5)
Constant	175.37*** (19.84)	198.56*** (20.14)	247.33*** (24.92)	1.13*** (0.11)	1.30*** (0.12)
Inclusion	30.86*** (4.40)	31.75*** (4.52)	29.30*** (4.37)	0.24*** (0.03)	0.25*** (0.03)
Ratings	11.23*** (2.40)	13.99*** (2.50)	10.92*** (2.38)	0.15*** (0.01)	0.16*** (0.01)
3-month int. rate	11.21*** (3.33)	13.21*** (3.41)		0.08*** (0.02)	0.10*** (0.02)
5-year int. rate	-55.25*** (7.11)	-62.81*** (7.23)		-0.37*** (0.04)	-0.42*** (0.04)
$\log(3\text{-month int. rate})$			28.21*** (7.35)		
$\log(5\text{-year int. rate})$			-208.85*** (22.33)		
Leverage	502.81*** (47.50)		495.17*** (47.28)	4.47*** (0.27)	
$\log(\text{Leverage})$		42.55*** (6.30)			0.49*** (0.04)
Index Returns	-231.69*** (83.43)	-231.25*** (85.64)	-199.87** (83.37)	-1.76*** (0.47)	-1.77*** (0.50)
Obs	1196	1196	1196	1196	1196
R^2	0.35	0.32	0.36	0.55	0.53

Table 2.5: Change in correlations after inclusion

This table shows the correlation between CDS spreads and CDX index before and after inclusion. The number pairs in parentheses are the standard deviations of ΔCDS and ΔCDX index, respectively. The fifth column shows the change in correlations. The last column indicates whether the standard deviation of CDS spreads after inclusion is much higher than before inclusion (≥ 3 times), in which case it may not make sense to compare correlations.

Trading Symbol	Inclusion Date	Correlation Before	Correlation After	Change(After -Before)	High Volatility After?
HOT	9/21/2006	0.57 (7.18 , 2.48)	-0.07 (9.52 , 1.35)	-0.64	
OLN	9/21/2006	0.06 (5.31 , 2.48)	0.40 (8.20 , 1.35)	0.34	
BSX	3/21/2007	0.13 (8.52 , 1.37)	0.33 (33.56 , 1.37)	0.20	
JCP	3/21/2007	0.34 (4.02 , 1.37)	0.81 (13.05 , 1.37)	0.47	yes
RDN	3/21/2007	0.89 (7.23 , 1.37)	0.52 (146.33 , 1.37)	-0.37	
DRI	9/21/2007	0.31 (5.61 , 8.13)	0.42 (23.47 , 10.87)	0.11	
FO	9/21/2007	0.49 (5.77 , 8.13)	0.51 (12.03 , 10.87)	0.02	yes
GCI	9/21/2007	0.61 (5.27 , 8.13)	0.44 (22.09 , 10.87)	-0.18	
HD	9/21/2007	0.79 (4.65 , 8.13)	0.60 (13.59 , 10.87)	-0.19	
LIZ	9/21/2007	0.37 (6.64 , 8.13)	0.31 (32.48 , 10.87)	-0.06	yes

Table 2.6: Change in correlations after exclusion

This table shows the correlation between CDS spreads and CDX index before and after exclusion. The number pairs in parentheses are the standard deviations of ΔCDS and ΔCDX index, respectively. The fifth column shows the change in correlations. The last column shows the ratio of standard deviation of CDS spreads after to before exclusion.

Trading Symbol	Exclusion Date	Correlation Before	Correlation After	Change(After -Before)	Stdev After / Stdev Before
WEN	9/21/2006	0.55 (6.47 , 2.48)	0.39 (6.27 , 1.35)	-0.16	0.97
GPS	3/21/2007	0.57 (11.36 , 1.37)	0.58 (16.62 , 1.37)	0.01	1.46
RSH	3/21/2007	0.07 (12.84 , 1.37)	0.75 (22.59 , 1.37)	0.69	1.76
BSX	9/21/2007	0.56 (34.21 , 8.13)	0.77 (19.83 , 10.87)	0.21	0.58
OLN	9/21/2007	0.78 (15.34 , 8.13)	0.67 (17.22 , 10.87)	-0.11	1.12
TIN	9/21/2007	0.89 (19.81 , 8.13)	0.71 (20.38 , 10.87)	-0.19	1.03

Chapter 3

The Recovery Factor in Credit Default Swaps

3.1 Introduction

Credit derivatives provide insurance against the risk of a default by a particular company or country on its bonds. The price of credit derivatives depends on two main factors: the risk-neutral probability of default and the risk-neutral recovery rates. Research on credit derivatives focuses mostly on the first factor, while regarding the second factor as a constant at about 0.4-0.5 of the face value. However, recent research has found that the realized recovery factor is not constant, and exhibits fluctuations as large as 0.2-0.7. Moreover, the realized recovery rates exhibit a high correlation with the probability of default, i.e., recovery rates are low when default rates are high, and vice versa.

Recovery rates are hard to capture in the pricing model. First of all,

defaults are rare and thus recoveries are also rare. As a consequence, we do not know much about both default probabilities and recovery rates and what the right statistics should be. Models for defaults have been developed and enable researchers to use CDS prices to back out implied default probabilities instead of using the historical default data. CDS spreads are also more informative about default than bond prices, since the bond market is much more complicated and less liquid than the credit market. Longstaff et al. (2005) uses CDS information to measure the size of default in corporate bond spreads, indicating the belief that CDS prices measure "pure" default risk. However, if recovery rates are important in the price and we do not understand this factor, then the credit pricing model cannot be complete, and the implied default probabilities may not be correct.

The information from CDS spreads is crucial for risk managers. Even though CDSs are priced in the risk-neutral measure and the implied default probabilities are in the risk-neutral world, the spreads are informative about the market view of default probabilities. Risk managers can use this information to construct their portfolios even though all they care about is the "physical" default. Implied recovery rates from CDS spreads can play the same role to the risk managers – they are informative about the market view of recovery rates.

In this paper we seek to understand and parametrize the risk-neutral recovery rates in CDS. We first detect the recovery factor in CDS using the unique characteristic of ex post recovery rates. Having learned that recovery rates are stochastic with a particular pattern, we proceed to modify the credit pricing model to capture this empirical fact. Along the process, we also

contribute to the empirical determinants of CDS spreads. The theoretical factors and other control variables can explain up to 90% (65% without fixed effects) of the CDS spreads in panel data, confirming that our pricing model works well with this seemingly complicated product.

Empirically, it has been difficult to separate the effect of risk-neutral recovery rates from the risk-neutral probability of default. The main difficulty is that both factors, which are stochastic with high correlations, are multiplied together to form a price. Moreover, it is not clear whether the recovery factor is incorporated into the price of credit derivatives ex ante. Empirical evidence that the recovery rates exhibit high correlations with default rates is found "after" the default occurs. Does the market incorporate this ex post result into the ex ante price before default? If so, do ex ante (risk-neutral) recovery rates exhibit the same characteristics as the ex post (physical) recovery rates?

We establish the fact that recovery rates are priced into the credit derivatives ex ante, in addition to the effect of the probability of default. The key to distinguish the effect of recovery rates from the probability of default is industry characteristics and industry distress factors. Following Shleifer and Vishny (1992) and Acharya et al. (2007), we know that recovery rates depend on industry characteristics and industry distress. The main drivers for recoveries are asset-specific recovery rates and the fire-sales effect. These two factors determine the recovery rates, in addition to effect of the probability of default. Using these two factors, and controlling for other factors that affect the CDS spreads, including the probability of default, we can detect the recovery factor in credit derivative prices.

This paper builds on the empirical findings on realized recovery rates. The three main papers in this area are Altman et al. (2005), Shleifer and Vishny (1992), and Acharya et al. (2007). Altman et al. (2005) finds that realized recovery rates are negatively correlated with the aggregate default rates. They explain that recovery rates are basically a function of supply and demand of the defaulted securities. Shleifer and Vishny (1992) explores the determinants of the liquidation values of assets. We quote from the abstract: "When a firm in financial distress needs to sell assets, its industry peers are likely to be experiencing problems themselves, leading to asset sales at prices below value in best use. Such illiquidity makes assets cheap in bad times, and so ex ante is a significant private cost of leverage". The main insight for recovery rates comes from the following observations:

- **Asset Redeployability:** Assets which are redeployable have high liquidation values. If they are managed improperly, the creditors can take the assets away and redeploy them. In general, asset redeployability depends on the industry the firms are in. For example, commercial land can be used for many different purposes, while insurance contracts may not be used for any other purposes. We call this factor industry characteristics.
- **Asset Specificity:** Assets which are specialized can be best used by specialists in the same industry. For example, aircrafts are best used by airline companies, but not retailers. When industry buyers cannot buy the assets and industry outsiders face significant costs of acquiring and managing the assets, assets in liquidation will have low recovery rates. This scenario can happen when the industry is in distress. Thus,

we call this factor industry distress.

Apparently the determinants of recovery rates in Shleifer and Vishny (1992) are different from Altman et al. (2005). Acharya et al. (2007) shows empirically that creditors of defaulted firms recover significantly lower amounts when the industry of the defaulted firms is in distress. The authors investigate whether the result is purely an economic-downturn effect or also a fire-sales effect along the lines of Shleifer and Vishny. They find that the fire-sales effect is also at work; creditors recover less if the industry is in distress and non-defaulted firms in the industry are illiquid, particularly if the assets are specific to the industry.

They study realized (ex post) recovery rates and find that industry characteristics and industry distress have a significant role in determining the recovery rates, even after controlling for the ex ante probability of default. They conclude that in addition to (and possibly instead of) the illiquidity in the financial market for trading defaulted instruments, illiquidity in the market for the sale of real assets is important in understanding creditor recoveries.

With these results on the recovery rates, we should expect to see similar effects of the risk-neutral recovery rates on CDS spreads. In particular, we should detect the effect of industry characteristics and industry distress on the CDS spreads. This will be our main empirical test.

Few papers explore the implied (risk-neutral) recovery rates in fixed income products. The closest ones are Bakshi et al. (2006), Le (2007) and Pan and Singleton (2008). We mention each of them here and specify how our paper is different.

Bakshi et al. (2006) explores the role of recovery in default risk models. They find that the recovery rate as a fraction of face value (RFV) fits the data the best. They specify the theoretical price for defaultable bonds, and then extract the implied recovery rates from the bond data. Their empirical results show that there is a relationship between risk-neutral expected default rates and risk-neutral expected recovery rates in bond prices. On average, a 4% worsening in the risk-neutral hazard rate is associated with a 1% decline in risk-neutral recovery rates. Our paper is different from their paper in a few aspects. First, we use credit derivative data which are relatively new compared to bond data. Most of the industrial credit models still assume that recovery rates are constant. Although there is a relationship between expected default rates and recovery rates in the bond market, it is not clear, or not established, that market participants in the credit market take this relationship into account. Second, in their model, the risk-neutral recovery rates depend on only the risk-neutral probability of default. In our paper, we specify that the risk-neutral recovery rates depend on industry characteristics and industry distress, which turn out to correlate with the probability of default. Third, they use a reduced-form model while we use a structural model for default risk. The probability of default in a structural model is clearly defined with a closed-form solution. In a reduced-form model, the probability of default is an (unobserved) Poisson jump process. This difference may affect how we estimate the probability of default, and correspondingly, the recovery rates, which are highly correlated.

Le (2007) proposes a framework to separate risk-neutral recovery rates and risk-neutral probability of default. The key is to use equity derivatives

to back out the probability of default, and then use that probability to price a CDS of the same firm. Given that we know the probability of default, the remaining factor that drives CDS spreads must be the recovery rates. The paper also finds that the probability of default and expected recovery rates are positively correlated. The motivation of his paper is similar to ours. However, our paper is more focused on which factor determines the expected recovery rates, rather than the general framework to separate the recovery rates and the probability of default. Since we know the factors that drive ex post (physical) recovery rates, we can use those factors to detect the effect of ex ante (risk-neutral) recovery rates. With this extra knowledge, we only need the CDS spreads and not equity options prices. The ultimate goal of our paper is to parametrize the stochastic recovery rates to price credit derivatives, and so our approach is more straightforward and less cumbersome but possibly lacks the generality as in Le (2007). Moreover, the recovery rates are stochastic in time series while Le (2007) only explores the determinants of recovery rates in cross section.

Pan and Singleton (2008) explores the nature of default arrival and recovery rates implicit in the term structure of sovereign CDS spreads. They use time series of CDS to estimate the default arrival and recovery rates. However, the paper still sets recovery rates as a constant, arguing that this is the industry standard and recoveries tend to be constant for sovereign debts. Our paper uses corporate CDS data instead of sovereign CDS data. Moreover, the main point of our paper is that the ex ante recovery rates are stochastic, thus setting recovery rates as a constant is not an option. Finally, our paper uses a structural model while their paper uses a reduced-form model.

To be consistent with the empirical findings, we incorporate the stochastic recovery rates into the credit risk model. We use the first-passage-time structural model to calculate the probability of default. We parametrize the recovery rate so that it depends on industry characteristics and industry distress, as suggested by the data. This step links the unobserved ex ante (risk-neutral) recovery rates to observable parameters such as industry index. We then calibrate the model to the data to find the value of parameters for risk-neutral recovery rates. The result is that the risk-neutral recovery rate will be lower by 20% if the industry is in distress during the firm's default. The model can be used to learn about the expected recovery rates across business cycles and it is possible to get the number different from 20% if calibrated to the data in different periods.

The chapter proceeds as follows. Section 3.2 describes the CDS pricing model. Section 3.3 describes how to calculate the probability of default. Section 3.4 describes the CDS data. Section 3.5 is an empirical analysis of the implied recovery factor in CDS. Section 3.6 is a robustness check. Section 3.7 describes the theoretical model we derive to capture the empirical findings. We then conclude in the final section.

3.2 CDS Pricing Model

In this section we introduce the CDS pricing model that will incorporate all the factors needed to be considered in the empirical analysis. A Credit Default Swap (CDS) is a contract that provides insurance against the risk of a default by a particular company. The company is known as the reference

entity and a default by the company is known as a credit event. The buyer of this contract obtains the right to sell bonds issued by the reference entity for their face value when a credit event occurs. On the other hand, the CDS seller agrees to buy the bonds for their face value when a credit event occurs, thus bearing the risk of default.

The CDS buyer makes periodic payments to the seller until the end of the life of the CDS or until a credit event occurs. The settlement, in the event of a default, involves either physical delivery of the bonds or a cash payment. The observable quantity in the market is the payment by the CDS buyer, also known as CDS spreads. The CDS spreads in a liquid market reflect the fair price for default insurance, i.e., the spreads must make the expected value of the buyer's periodic payments equal to the expected value of the seller's losses in case of default.

In a no-arbitrage model, the CDS spreads should make the present value of payments from the buyer (fixed leg) equal to the present value of losses by the seller (contingent leg). Let S be the CDS spread and R the risk-neutral expected recovery rate if default occurs. Let $Q(t)$ be the risk-neutral cumulative probability of default up to time t , thus the risk-neutral probability of survival up to time $t = 1 - Q(t)$. Let d be the accrual days between payment dates, for example, if the payments are made quarterly, then $d = 0.25$. Let $D(t)$ be the discount factor. Then

$$PV[\text{fixed leg}] = \sum_{i=1}^N D(t_i)(1 - Q(t_i))Sd_i$$

$$PV[\text{contingent leg}] = (1 - R) \sum_{i=1}^N D(t_i)(Q(t_i) - Q(t_{i-1}))$$

The first equation represents the expected payment from the buyer, by mul-

tipling payments with the probability of survival. The second equation represents the expected payment from the seller, by multiplying the payment in case of default with the default probability. By equating the two present values, we can represent S as:

$$S = \frac{(1 - R) \sum_{i=1}^N D(t_i)(Q(t_i) - Q(t_{i-1}))}{\sum_{i=1}^N D(t_i)(1 - Q(t_i))d_i} \quad (3.1)$$

The above equation is for discrete time. The continuous-time model uses the same intuition. We take CDS spread formula, given for the contract starting from time 0 to T , from Hull, Predescu and White (2009):

$$S = \frac{(1 - \hat{R})(1 + a^*) \int_0^T q(u)v(u)du}{\int_0^T q(u)[h(u) + e(u)]du + (1 - \int_0^T q(u)du)h(T)} \quad (3.2)$$

where

T : The period of the contract

$q(u)$: Probability density function (pdf) of default at time u in a risk-neutral world

\hat{R} : Expected recovery rate on the reference obligation in a risk-neutral world

$h(u)$: Present value of payments at the rate \$1 per year on payment dates between time zero and time u

$e(u)$: Present value of accrual payment at time u equal to $u - u^*$ where u^* is the payment date immediately preceding time u

$v(u)$: Present value of \$1 received at time u

a^* : Average value of accrued interest rate on the reference obligation for the

period 0 to T

We will use the discrete-time model to guide our empirical analysis, but we will use the continuous-time model when we modify the recovery rate function in the theoretical section. From (3.1), we can see that the factors that affect CDS spreads are the risk-neutral probability of default, the risk-neutral recovery rate, and the interest rate (discount factor). In this paper, for simplicity, we assume the interest rate is constant when investors price the financial instruments. We can then write $D(t) = e^{-rt}$.

3.3 The Probability of Default

There are two main approaches to model the probability of default: structural models and reduced-form models. A structural model characterizes defaults as asset values falling below default barriers. A reduced-form model abstracts away from the fundamentals of the firms and characterizes defaults as a Poisson process. The advantage of a structural model is that it is economically intuitive, while the advantage of a reduced-form model is that it is easier to fit the data to the "unobserved" default intensity. In this paper we only use a structural model approach. The choice here is not only because of the economic intuition, but also because we will incorporate the results from corporate finance into the recovery factor. In corporate finance models, assets and debts have "physical" meaning, not just a mathematical concept, and so we do not want to abstract away from the physical model. We would like to have a unified view of default processes and recoveries in a

single model.

In a structural model, firms default when they cannot fulfill financial obligations. In mathematical terms, the firm defaults when its value falls below the debt barrier. The first generation model by Merton calculates the probability of default at the date of debt maturity, for example 1 year from now. The firm can only default at the date of debt maturity. The next generation of model, pioneered by Black and Cox (1976), characterizes default as the first passage time of the firm value across the debt barrier. In this model firms can default any time between now and the date of debt maturity. Mathematically it is much harder to calculate the first passage time density as in the Black and Cox model than to calculate the probability of default at maturity date as in the Merton model. In empirical analysis section we use the probability of default from the Merton model, which was modified by Bharath and Shumway (2008). We briefly describe the calculation method here.

In the risk-neutral measure, the (unobserved) value of a firm follows a geometric Brownian motion

$$dV = rVdt + \sigma_V V dW^Q \quad (3.3)$$

which is a standard setup and r is the risk-free rate. The firm issues a discounted bond maturing at time T . Under this assumption, the equity of the firm is a call option on the underlying value with a strike price equal to the face value of the firm's debt and a time-to-maturity of T . The value of equity as a function of the unobserved total value of a firm can be described by the Black-Scholes-Merton formula.

Specifically, Merton shows that the equity value of a firm satisfies

$$E = VN(d_1) - e^{-rT}FN(d_2) \quad (3.4)$$

where E is the market value of the firm's equity, F is the face value of debt, $N(\cdot)$ is the cumulative standard normal distribution function, d_1 is calculated from

$$d_1 = \frac{\ln(V/F) + (r + \frac{1}{2}\sigma_V^2 T)}{\sigma_V \sqrt{T}} \quad (3.5)$$

and $d_2 = d_1 - \sigma_V \sqrt{T}$

The Merton model links the unobserved total value of the firm to the observed equity value. Moreover, with a closed-form expression, we can link the volatility of the firm to the volatility of the equity. It follows from Ito's lemma that

$$\begin{aligned} \sigma_E &= \left(\frac{V}{E} \right) \frac{\partial E}{\partial V} \sigma_V \\ &= \left(\frac{V}{E} \right) N(d_1) \sigma_V \end{aligned} \quad (3.6)$$

where the second line comes from the fact that $\frac{\partial E}{\partial V} = N(d_1)$ from the partial derivative of Eq(3.4).

From Eq(3.4) and Eq(3.6), we have two equations and two unknowns (V and σ_V). Thus, we can solve for the unobserved variables V and σ_V , given the value of E and σ_E .

Once we solve for V and σ_V , risk-neutral the distance to default (DD) can be calculated as

$$DD = \frac{\ln(V/F) + (r - \frac{1}{2}\sigma_V^2 T)}{\sigma_V \sqrt{T}} \quad (3.7)$$

and then the risk-neutral probability of default (PD) is given by

$$PD = N(-DD) \quad (3.8)$$

The probability of default here characterizes the probability that the firm value will fall below the face value of debt at maturity.

While the mentioned method is straightforward, it involves solving simultaneous equations for every observation. In the paper by Bharath and Shumway (2008), the authors explore an alternative way to calculate the probability of default. The method is based on the same economic intuition about the default process, but does not require solving simultaneous equations. We briefly describe their method here:

To begin, they approximate the market value of each firm's debt with the face value of its debt,

$$D_{BS} = F \quad (3.9)$$

Where BS stands for Bharath-Shumway. The volatility of the debt is correlated with the equity volatility

$$\sigma_{D,BS} = 0.05 + 0.25 * \sigma_E \quad (3.10)$$

The 5 percent represents the term structure volatility and the 25 percent times equity volatility represents the volatility associated with the default risk. Thus, the approximation of the total volatility of the firm is calculated by

$$\begin{aligned} \sigma_{V,BS} &= \frac{E}{E + D_{BS}} \sigma_E + \frac{D_{BS}}{E + D_{BS}} \sigma_{D,BS} \\ &= \frac{E}{E + F} \sigma_E + \frac{F}{E + F} (0.05 + 0.25 * \sigma_E) \end{aligned} \quad (3.11)$$

This approximation captures the same information that is captured by the Merton model, but without having to solve simultaneous equations. The naive distance to default is then given by:

$$DD_{BS} = \frac{\ln((E + F)/F) + (r - \frac{1}{2}\sigma_{V,BS}^2)}{\sigma_{V,BS}\sqrt{T}} \quad (3.12)$$

The DD_{BS} is easy to compute and also represents the same information as the Merton DD . Finally the probability of default is given by

$$PD_{BS} = N(-DD_{BS}) \quad (3.13)$$

In section 4.5 of Bharath and Shumway (2008), the authors regress the log of CDS spreads with the log of PD_{BS} and compare the result with the regression with the log of Merton PD . It should be noted that they use 1-year PD while the CDS contract is 5-year. The reason is that 1-year PD is highly correlated with 5-year PD , and the log specification of this regression makes the intercept reflect the average level of the probability. The regression results show that both variables are statistically significant, but the R^2 of $\log(PD_{BS})$ is higher than $\log(\text{Merton } PD)$. Moreover, when adding both PD_{BS} and Merton PD in the same regression, the statistical significance of the Merton PD is driven out by the PD_{BS} . They conclude that the functional form of the probability of default is more important than the solution procedure.

Due to the result of this paper and relative ease of computation, we will use the PD_{BS} in our empirical analysis instead of the original Merton PD . We have explained the determinants and calculation of the recovery rates and the probability of default. We are now ready for the empirical analysis of the CDS spreads.

3.4 Data

We use CDS data from Credit Market Analysis (CMA), acquired by CME on March 25, 2008. The data are daily ranging from January 2004 to May 2008. For this paper we use only 5-year CDS data because they are the most common and most liquid. From daily data, we change the interval to monthly, because it is known that the CDS spreads have high autocorrelations, possibly because of illiquidity. We use the spreads at the end of month as monthly data. We match firms in our CDS dataset with CRSP database using Ticker symbols. The summary statistics of the CDS data are presented in Table 3.9.

To detect the effect of recovery rates, we use industry dummy variables and industry distress indicators. The industry dummy variables are assigned according to the firm's Siccocode. We divide firms into 10 industries based on the Siccodes on Kenneth French's website. The percentage of firms in each industry in our dataset is shown in Table 3.2.

The industry distress indicators need to signify the state of the industry. According to Acharya et al. (2007), there are various ways to define industry distress. The first is that the median stock return for the industry of the defaulting firm falls below -30% annually. This accounts for 9% of the sample data. The second is that one-year or two-year median sales growth for the industry is negative. The third is that the average credit rating of other firms in the industry is below investment grade. The three proxies give similar results in their paper and the first proxy is used for all subsequent analyses.

Following the first proxy of industry distress, we use the 10th percentile

as a cutoff for distress. If the industry return for that month is less than the 10th percentile, then industry distress indicator is 1, otherwise it is 0. We use the 10th percentile cutoff here to be consistent with the literature but the result is also robust to other cutoff levels. The cutoff comes from the historical data of industry returns for each industry during the 10-year period of 1994-2003. The distress cutoff for each industry is shown in Table 3.3.

To control for the probability of default, we calculate the probability of default using PD_{BS} as described in the last section. We use the equity data from CRSP and the debt data from Compustat. Following Bharath and Shumway (2008), the face value of debt (F) is calculated as (short-term debt + 0.5*long-term debt). The equity volatility (σ_E) is also calculated from monthly returns from CRSP, and then adjusted to annual scale. We use the risk-free rate $r = 2.5\%$ for the drift in the risk-neutral measure. We want to fix the interest rate effect in the probability of default. This assumption may be relaxed and the main result still goes through.

Another important proxy for the probability of default is the firm's ratings. In theoretical models, PD should be sufficient statistics that capture all the information of default probabilities. However, in the data, ratings are significant in explaining credit spreads even after controlling for PD (Aunon-Nerin et al. (2002)). Even though ratings reflect physical instead of risk-neutral probability of default, the two must be highly correlated, and we can use one as a proxy for another in the regression analysis. Thus, in the regression we include the probability of default as reflected by the firm's ratings. We use S&P ratings available on Compustat. We then convert ratings

into the equivalent probability of default using Moody's corporate idealized 5-year cumulative probability of default rates.

Other factors that may affect CDS spreads include interest rates as a discount factor. We use Treasury bill rates of maturity 3-month and 5-year as a control. These are the shortest and longest maturity interest rates that may affect CDS spreads. Investors may use different interest rates to discount cash flows, but they should be in the range of these short and long term rates, or their linear combination. In the empirical analysis, we only include these two rates. The data are also from CRSP.

3.5 Empirical Results

The main hypothesis to test here is whether recovery rates are incorporated into the CDS spreads, *ex ante*. The results from Acharya et al. indicate that industry characteristics and industry distress determine physical recovery rates, *ex post*, in addition to the risk-neutral probability of default. We will then look for the significance of industry dummy variables and industry distress indicators in the regression with CDS data. The result is shown in Table 3.4.

The first column shows the full regression while the second column shows the regression without the industry condition (only the probability of default and interest rates). The R^2 in the second column decreases from the first column by about 0.034 or 3.4%. The industry condition can explain about 3.4% of the CDS spreads. Most of the variations in CDS spreads indeed come from the probability of default. However, the industry condition is also

highly significant and the economic magnitude is big, especially for *Distress*. We now focus on the significance of the coefficients in the first column.

After controlling for the probability of default and interest rates, the industry dummy variables and industry distress indicators are still statistically significant. The variables $I1 - I9$ are industry dummy variables, where $I1 = 1$ if the firm is in industry number 1 and 0 otherwise, and similarly for $I2$ to $I9$. Note that we cannot include $I10$ in the regression, otherwise we will get a linearly dependent vector of $I10$ with $I1$ to $I9$. The industry effect of $I10$ is absorbed into the intercept in the regression. All the industry dummy variables are significant. This indicates that the industry-specific asset is an important factor to determine the expected recovery rates in CDS spreads. However, in this paper we put more focus on the *Distress* factor, which is time-varying. The industry dummy can be regarded as a constant which can vary by industry but will not explain the time-varying patterns of recovery rates over time.

The variables *Distress* and $lag(Distress)$ stand for the industry distress indicator and its 1-month lag. Both variables are significant and the sign is positive as expected. If the industry is in distress, then the recovery rates will be lower because it is hard to sell assets to non-defaulted firms in the same industry. Thus, fire-sales effect, which will occur when the industry is in distress, is also priced into the CDS spreads. The lag of *Distress* is included in the regression because we doubt that the CDS market may not be liquid and thus the information of *Distress* may take time to be incorporated into the price. Indeed, the lag of *Distress* is also significant in the regression. We will discuss more about illiquidity effect in the robustness check section.

The magnitude of *Distress* is 0.4, which is very big in a log scale. This translates to about $e^{0.4} - 1 = 0.5$ or 50% decline in expected risk-neutral recovery rates if the industry is in distress. This extreme magnitude can have many interpretations. First, it can mean that investors are very risk-averse about recovery rates and thus a sharp decline in expected recoveries in distress. Second, it can mean that our Distress condition is too extreme and thus the resulting magnitude is too high. Third, it can mean that Distress is a proxy for some other factors as well, for example illiquidity and time-varying risk premium. We will see that the third explanation is the most plausible and we provide evidence in the robustness check section.

Probability of Default (*PD*) and Interest Rates are included as control variables in the regression. With control variables, we make sure that industry characteristics and industry distress variables are significant because of recovery rates, not the probability of default or other factors. Note that *PD* and interest rates are also significant in the regression as we expect from the theoretical formula. However, we focus only on the recovery rates and do not go into details about these two factors in this paper.

Since this is a panel regression, we should also control for the firm fixed effect and time fixed effect. We do the same regression with firm and time fixed effects and the result is shown in Table 3.5.

The first column is the same as the first column in the previous table. In the second column we add firm fixed effect and yearly time fixed effect. In the third column we add firm fixed effect and monthly time fixed effect.

As we can see from the second column, some of the industry characteristics ($I1 - I9$) lose statistical significance. This result is as expected. When

we control for finer information (firm-level), then the cruder information (industry-level) is no longer significant. The coefficients for *Distress* and $\text{lag}(\text{Distress})$ are still significant although the magnitude is slightly smaller than the first column. Note that the R^2 jumps by almost 20% when we include firm and time fixed effect.

In the third column, we see the same result for industry characteristics ($I1 - I9$). However, *Distress* and $\text{lag}(\text{Distress})$ have much lower magnitude than before, although they are still highly significant. When we control for time fixed effect at the monthly level, then this information is highly correlated with *Distress*. Industry conditions are highly correlated with the economy and month fixed effect may be a proxy for the state of the economy for a given month. Nevertheless, after controlling for month fixed effect, *Distress* is still highly significant. This means that the industry condition does affect the recovery rates and this factor is not purely driven by the macro economy.

3.6 Robustness Check

The results from the last section indicate that industry characteristics and industry distress are significant in determining CDS spreads. In other words, the recovery factor is priced into CDS spreads ex ante. In this section, we check for the robustness of the result. In particular, we check whether other factors may affect CDS spreads and make industry characteristics and industry distress no more statistically significant in the empirical analysis.

The first factor is the state of the economy. Obviously the state of the

economy will affect the probability of default, recovery rates and interest rates and thus the CDS spreads. It can also affect the level of risk-aversion of investors and change the probability in the risk-neutral measure. We control for the state of the economy by using the S&P500 index returns from CRSP.

The second factor is the state of the industry. Acharya et al. (2007) indicate that the continuous industry returns have no effect on recovery rates ex post, only the industry distress does. However, ex ante, industry returns may affect the expected recovery rates. Moreover, the industry returns may affect the probability of default. Thus, the industry returns can affect the CDS spreads and we include them as a control variable. We use industry returns data from Kenneth French's website.

The third factor is illiquidity. In a liquid market, the CDS spreads should reflect a fair price. However, in an illiquid market, it can be the case that CDS spreads are higher because no one is willing to provide credit protection. Illiquidity usually happens during the distressed period. Also, different industries may have different liquidity for CDS protection. Thus, our results in the previous section may reflect market illiquidity factor, not the recovery factor in CDS spreads. For robustness check, we control for the illiquidity factor.

In the literature it is still not clear how to measure illiquidity. The main indicator of illiquidity is bid-ask spreads. However, it is not clear what functional form of illiquidity affects the price. Here we control for illiquidity by using the bid-ask spreads of the CDS prices. As for the functional form, we simply run regressions on multiple functional forms of bid-ask spreads and

report the one with the highest R^2 result. It turns out that the log scale results in the highest R^2 for the regression.

In theory we should include fixed effects in all regressions. However, as we have seen from Table 3.5, including firm fixed effect will naturally dry out the industry characteristics effect. Also, including month fixed effect will naturally dry out the industry distress effect. Thus, in this section, we first run a regression without fixed effects in one table and then run a regression with fixed effects in the next table.

Table 3.6 shows the result of the regression with control variables but without fixed effects. The first column shows the original regression as in Table 3.4. The second column shows the regression with additional S&P500 index returns as a control for the state of the economy. The third column shows the regression with additional industry returns as a control for the state of the industry. The fourth column shows the regression with $\log(\text{Bid-Ask spreads})$ as a control for illiquidity. The fifth column shows the regression with all the control variables.

When we add index returns as a control in the second column, all the regression coefficients are the same except for *Distress*, whose coefficient decreases from 0.40 to 0.31 but is still significant. The coefficient for index returns itself is also highly significant with a negative sign as expected. This means that the state of the economy can influence the CDS spreads in addition to the factors in the original regression. More interestingly, the state of the economy can affect the recovery factor that used to be captured by the industry distress factor. However, since *Distress* is still significant, we see that industry distress still determines the recovery rates after controlling

for the economy. The $\text{lag}(\text{Distress})$ factor remains highly significant with high magnitude. Perhaps $\text{lag}(\text{Distress})$ is even a better proxy for industry distress, because the CDS market may take some time to incorporate this information into the price. We will see this phenomenon more in the later regressions.

When we add industry returns as a control in the third column, all the regression coefficients are the same except for *Distress*, whose coefficient decreases from 0.40 to 0.28 but is still highly significant. The coefficient for industry returns itself is also highly significant with a negative sign as expected. Similar to index returns, industry returns can affect the recovery factor that used to be captured by the industry distress factor, even more so than index returns. However, since *Distress* is still significant, the industry distress factor still determines the recovery rates after controlling for the state of the industry.

When we add illiquidity as a control in the fourth column, most of the previous coefficients including the $I1 - I9$ industry characteristics variables remain significant except for $I2$. The coefficients also change from the original regression. The coefficient for illiquidity itself is highly significant with a positive sign as expected. Thus, illiquidity does affect the CDS spreads that used to be captured by the industry characteristics. The more drastic effect is on the *Distress* and $\text{lag}(\text{Distress})$ factors. After controlling for illiquidity, the coefficients for *Distress* and $\text{lag}(\text{Distress})$ decrease from 0.40 to 0.20 and 0.39 to 0.19, respectively. This means that during the period of illiquid market, CDS spreads are higher because it is hard to find a seller. Moreover, the market is usually illiquid when the industry is also in distress,

which is also when the recovery rates are expected to be lower. The effect of recovery rates in distress, and also the effect of illiquidity on CDS spreads, were captured by the *Distress* and $\text{lag}(\text{Distress})$ factors in the original regression. However, when we include illiquidity factor as a control variable in this regression, we separate the effects of industry distress and illiquidity. As a result, we see the coefficients for *Distress* and $\text{lag}(\text{Distress})$ decrease significantly. However, since both factors are still highly significant, the industry distress factor still determines the recovery rates after controlling for illiquidity.

Finally we include all control variables in the fifth column. The coefficients of I1-I9 are very similar to the fourth column. The coefficient for $\text{lag}(\text{Distress})$ stays the same as in the fourth column, but the coefficient for *Distress* decreases from 0.20 to 0.13 and becomes insignificant at 10% level. This means that the illiquidity factor affects CDS spreads in a similar way as industry characteristics and industry distress, more so than index returns and industry returns. The coefficient of *Distress* decreases from the fourth column because *Distress* is also affected by index returns and industry returns, as shown in the second and third column.

The magnitude of *Distress* and $\text{lag}(\text{Distress})$ in the last column is more realistic. The magnitude of 0.13 and 0.19 in the log scale translates to $e^{0.13} - 1 = 0.14$ and $e^{0.19} - 1 = 0.21$, or 14% and 21% decrease in risk-neutral recovery rates. This magnitude is also consistent with the result from our theoretical section. We also see that *Distress* itself is not significant but $\text{lag}(\text{Distress})$ is still significant. This can mean that the information about industry distress takes some time to be realized in the market price because the CDS market

is not perfectly liquid.

In sum, after controlling for all other factors that may affect CDS spreads, the industry characteristics and industry distress factors are still highly significant. In particular, the significance of the industry distress indicator confirms that the stochastic recovery rates are priced into CDS spreads, *ex ante*.

Now we add in the firm and time fixed effects in the regression. The result is reported in Table 3.7.

The first column of Table 3.7 shows the regression with no fixed effect, the same as in the last column of Table 3.6. The second column shows the regression with only firm fixed effect. The third and fourth columns show the regression with firm-year and firm-month fixed effect. Note that we again see a jump in R^2 from the first column to the next three columns. Fixed effects indeed play a significant role in explaining CDS spreads. The R^2 is almost 90%. This is quite interesting. As complicated as CDS cash flows are, only a handful of factors can explain almost 90% of all the spreads.

In the second column we see the same effect as before. After controlling for firm fixed effect, some of the industry characteristics lose statistical significance. However, $\text{lag}(\text{Distress})$ is still highly significant with roughly the same magnitude while *Distress* becomes significant at 10% level. The result in the third column is similar to the second column. In the fourth column, after controlling for the time fixed effect at monthly level, we see the magnitude of *Distress* and $\text{lag}(\text{Distress})$ decreases substantially. However, both coefficients are still highly significant. Note that the magnitude of *Distress* and $\text{lag}(\text{Distress})$ in the first three columns stays roughly at

0.2 which translates into about 20% decrease in risk-neutral recovery rates in distress. In the last column the magnitude decreases to 0.09 or 9% in linear scale. This can also mean that the month fixed effect absorbs too much information from the industry condition. Controlling for such fine information will naturally, or perhaps too severely, dry out the significance and magnitude of other variables. After all, putting in too many time fixed effects will drive away the effect of any time-varying variable which we want to investigate. Nevertheless, the coefficients are still significant and our main hypothesis still goes through.

3.6.1 Time-Varying Risk Premium

It has been shown that risk premium is also time-varying, higher in recessions and lower in booms. The risk premium is the difference of stochastic quantities in the physical and risk-neutral measure. In this case, our concern is that the risk premium changes the risk-neutral default probabilities even if the firm has the same distance-to-default. It is possible that the industry distress factor is just a proxy for bad times and its significance may just mean that in bad times the risk-neutral default probability is higher, but recovery rates are unchanged. With this concern, we try to control for time-varying risk premium in this section.

Before and During the Crisis

Risk premium should increase during the crisis, so as industry distress. Thus, industry distress may just be a proxy for an increase in risk-neutral default probabilities, and not stochastic recovery rates. In this subsection we split the

data into before and during the crisis period. The before-crisis period is from 2004-2006 and the during-crisis period is from 2007 to mid 2008. This way, industry distress condition is not just a proxy for a change in risk premium. We run the same regression as in Table 3.7 with firm and year fixed effects. We drop firms with the market cap more than 5% of the market (28 firms out of 332). These firms may have too strong effects on the industry distress dummy variable. In the previous regressions we also tried dropping these firms and the results did not change. The result is reported in Table 3.8. We only report the coefficients for default probabilities and recovery rates, i.e., $Distress$, $lag(Distress)$, $log(PD)$, and $log(PD \text{ from Ratings})$. We are not concerned with the significance of industry dummy or other control variables here.

Even after splitting the data into two subperiods, the coefficient for $lag(Distress)$ is still significant for both subperiods. Interestingly, before the crisis, only the $lag(Distress)$ is significant, and during the crisis, both $Distress$ and its lag are significant with higher magnitude. This gives us an opportunity to learn about behavioral credit market. During the boom, information about industry distress takes time to get incorporated into the price. During the crisis, investors pay particular attention to bad news, and thus industry distress information gets incorporated into prices very quickly.

Time-varying Risk-Neutral Probability of Default

Now we tackle the problem of time-varying risk-neutral probability of default in another way. We assume that risk premium can change over time and thus the risk-neutral probability of default can change even if the firm has the

same characteristics in the physical measure. We include the interaction term between $\log(PD)$ and $Year$ ($\log PD * Year_{2005}$, for example) and $\log(PD)$ from Ratings) and $Year$ in the regression. The idea is that there is an average effect of probability of default on CDS spreads. However, as time changes, the risk-neutral probability of default changes and this will reflect in the interaction term. The result is reported in Table 3.9. We do not report the coefficients for control variables.

As we can see from Table 3.9, $\text{lag}(Distress)$ is still highly significant after controlling for time-varying probability of default. $Distress$ has the right sign and reasonable magnitude but is not significant, which, again, may reflect how information is incorporated into the price in the CDS market. This result again confirms our hypothesis that risk-neutral recovery rates are indeed linked to the industry distress condition.

3.6.2 Using Implied Volatility

In the previous regressions we calculate the asset volatility using historical equity volatility as an input. This method may raise some concerns about the forward-looking nature of CDS spreads. Using historical volatility to calculate the probability of default may not be accurate since the probability of default needs to be forward-looking. This is not a problem if the volatility is a constant but empirical evidence seems to suggest otherwise. In particular, the significance of $Distress$ and $\text{lag}(Distress)$ may be a proxy for the changes in forward-looking volatilities, which, in turn, increase the probability of default. In this section we use option implied volatilities as an input instead of historical equity volatility. Implied volatilities are forward-looking and

should eliminate the concern.

We use implied volatility surface from OptionMetrics. We use 1-year at-the-money call option with $\delta = 50$. We then run the exact same regression as in Table 3.7 and report the result in Table 3.10. We report only the coefficients related to industry distress and probability of default, i.e., $Distress$, $lag(Distress)$, $log(PD \text{ from Implied Volatility})$ and $log(PD \text{ from ratings})$. We conclude that our main results still go through using forward-looking implied volatility instead of historical equity volatility.

3.6.3 Lag Effect

In the previous regressions, it turns out that $Distress$ and $lag(Distress)$ are highly significant. In some instances, however, $Distress$ is not significant but $lag(Distress)$ is significant instead. It is understandable that $Distress$ should be significant because the recovery rates are expected to be lower during the industry distress period. However, why is $lag(Distress)$ also significant, and even more persistent than $Distress$? Why does $Distress$ still affect CDS spreads even after a month? In this section we provide a couple of possible explanations:

1. CDS market is illiquid:

It is generally believed that the CDS market is more liquid than the bonds market. However, as we have also touched upon in this paper, the CDS market is not "perfectly" liquid. We have shown that illiquidity factor (bid-ask spreads) can affect the CDS spreads. If the market is not liquid, then it can take time for new information to be incorporated into the price of CDS. For example, when the industry is in distress,

the recovery rates are expected to be lower, thus CDS spreads higher. However, it may be hard to find a seller in such situation and for a buyer to agree to pay a higher price. The information of industry distress may not show in the CDS prices until some later time. Thus, it is possible that $lag(Distress)$ is significant in explaining the CDS spreads.

2. Price discovery occurs in the equity market and not in the CDS market:

We use industry returns information from the equity market to define industry distress. However, as has been shown in Hilscher et al. (2010), equity returns lead credit protection returns, while credit protection returns do not lead equity returns. The authors interpret their findings as evidence that informed traders are primarily active in the equity market. They state that the participants in CDS markets do not pay sufficient attention to equity returns. If equity returns lead CDS returns, then using the information in equity returns to explain CDS spreads will exhibit some lag effect. In the case of industry distress which is related to default, CDS market participants may pay attention to such event and thus the information is incorporated into CDS spreads as shown by the *Distress* variable. However, market participants may not pay sufficient attention to equity returns and thus some information about industry distress is slow to be incorporated into CDS spreads, as shown by the $lag(Distress)$ variable.

3.7 Modified Theoretical Model

In this section we modify the credit derivatives pricing model to incorporate the empirical results in the previous sections. In the last paragraph of Acharya et al. (2007), the authors suggest future researchers to derive a general equilibrium model which analyzes the risk premium arising from the industry distress effect. However, in this paper we are only concerned with the pricing model with no-arbitrage assumption. Our first concern is that the stochastic recovery rates may not even be incorporated into the empirical prices ex ante. If that is the case, then it would be of no use to modify the theoretical model to incorporate stochastic recovery rates. In the previous section, we have already established that ex ante (risk-neutral) recovery rates exhibit the same characteristics as the ex post (physical) recovery rates. The remaining task is to incorporate these results into the theoretical model. The model links the unobserved risk-neutral recovery rates to the observable industry index.

We start with the CDS pricing equation in (3.2). The important quantity in this equation is $q(u)$. In a structural model we characterize default as the first passage time of the firm value (V) across the barrier (B). We assume the barrier grows at the same rate as the expected growth rate of the firm, i.e., the firm has a constant expected leverage ratio in the risk-neutral measure. This assumption turns out to simplify much of the mathematics involved. Economically, this assumption is supported by recent papers; Almeida and Philippon (2007) argues that firms evaluate capital structure decisions in the risk-neutral measure; Collin-Dufresne and Goldstein (2001) finds that a structural model with mean-reverting leverage ratios is more consistent

with empirical findings. Considering the two papers together, we assume in the model that firms maintain a constant expected leverage ratio in the risk-neutral measure. The probability of default is equal to the first passage time distribution. The analytical formula is well-known as given by Harrison (1985) and Shreve (2004) and is also used by Zhou (2001). We give a basic setup and the result here while the proof can be found in Appendix A.

In the risk-neutral measure, we have

$$\frac{dV}{V} = rdu + \sigma dW^Q$$

Let $B(u)$ be the default barrier at time u . By our assumption,

$$B(u) = B(0)e^{(r - \frac{\sigma^2}{2})u}$$

With this setup, we get the default density (or first passage time density) in the risk-neutral measure:

$$q(m, u) = \frac{m}{u\sqrt{2\pi u}} e^{-\frac{m^2}{2u}}$$

where $m = \frac{\log(\frac{V(0)}{B(0)})}{\sigma}$. Note that m is equivalent to the distance-to-default at the beginning of the CDS contract. The density $q(m, u)$ here is indeed what we want for $q(u)$ in equation 3.2.

Proof: See Appendix A.1

Figure 1.1 shows the probability density of default, $q(m, u)$, for selected values of m corresponding to different credit ratings. Figure 1.2 shows the historical default density from S&P report for firms with different credit ratings. Note that the two figures show similar patterns of default for different rated firms. We can conclude that the theoretical formula of default density,

$q(m, u)$, can capture real-world default probabilities, and thus can be used to calculate CDS spreads.

3.7.1 The Recovery Model

Empirical findings suggest that recovery rates depend on the industry condition; if the industry is in distress when the firm defaults, then the recovery rates will be lower than in the normal time. This result also has support from theoretical papers such as Shleifer and Vishny (1992). The empirical results from Acharya et al. (2007) and Altman et al. (2005) confirm the theory ex post (after the default occurs) and in the physical measure. Our empirical results from CDS data confirm that the ex post expectation is also incorporated into the credit derivatives prices ex ante, in the risk-neutral measure. In this section we suggest a theoretical model for the recovery factor that can be used to price credit derivatives.

The key quantity is the expected risk-neutral recovery rate given default. Let $R(t)$ be the risk-neutral recovery rate at time t . Then

$$R(t) = a_1 - a_2 \mathbb{1}_{\{Distress(t)|Default(t)\}} \quad (3.14)$$

This indicates that the risk-neutral recovery rate during normal time is a_1 while the risk-neutral recovery rate during the distressed period is lower by a_2 . The value of a_1 and a_2 can also vary from industry to industry.

In the pricing model, investors take expectation of recovery rates to price credit derivatives. Note that this expectation is taken in the risk-neutral measure, so the probability and magnitude may not be the same as in the

physical measure.

$$E^Q[R(t)|Default(t)] = a_1 - a_2 P\{Distress(t)|Default(t)\} \quad (3.15)$$

The result in (3.15) is a straightforward expectation from (3.14), where we note that the expectation of an indicator function is the probability of the event. The key quantity in the recovery factor in CDS is then $P\{Distress(t)|Default(t)\}$

A Simple Model

To calculate $P\{Distress(t)|Default(t)\}$, we identify the dynamics of the firm's value, firm's barrier, industry index and industry barrier as follows:

$$\frac{dV}{V} = rdt + \sigma_V dW^Q \quad (3.16)$$

$$B(t) = B(0)e^{(r - \frac{\sigma_V^2}{2})t} \quad (3.17)$$

$$\frac{dI}{I} = rdt + \sigma_I dZ^Q \quad (3.18)$$

$$D(t) = D(0) \quad (3.19)$$

$$corr(dZ, dW) = \rho \quad (3.20)$$

The first two equations for the firm's dynamics are similar to the setup in the previous section when we calculate the probability of default. The third equation is the industry index dynamics. We assume that industry index grows similarly to the firm in the risk-neutral measure. The fourth equation is the industry barrier dynamics which we assume to be a constant. If industry index, $I(t)$, falls below the industry barrier, $D(t)$, then we say that the industry is in distress at time t . Finally, we assign ρ to be the correlation between the firm's dynamics and the industry dynamics.

Given the dynamics of the firm and the industry, we can find the probability that the industry is in distress when the firm defaults as

$$P \{Distress(t)|Default(t)\} = \begin{cases} N \left(\frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sqrt{t(1-\rho)}\sigma_I} + \frac{\sqrt{\rho}m}{\sqrt{t(1-\rho)}} \right) & \text{if } \rho \geq 0 \\ N \left(\frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sqrt{t(1-|\rho|)}\sigma_I} - \frac{\sqrt{|\rho|m}}{\sqrt{t(1-|\rho|)}} \right) & \text{if } \rho < 0 \end{cases}$$

Proof: See Appendix A.5

For shorthand, we call the above function p , i.e.,

$$P \{Distress(t)|Default(t)\} = p(m(0), D(0), I(0), \rho, t) \quad (3.21)$$

The function $p(m(0), D(0), I(0), \rho, t)$ requires input $m(0)$, $D(0)$, $I(0)$ and ρ which are known when pricing the CDS. The probability is characterized by these four numbers and the time of interest (t). The distance-to-default (m) is calculated from the DD_{BS} in the previous section. The industry index (I) can be observed and the correlation (ρ) can be estimated. The only quantity that we have not specified is $D(0)$ or the industry distress barrier at time 0. To the best of our knowledge, there is no standard way to define industry distress. We propose the following definition for industry distress which we will use to price CDS.

$$D(0) = c \left(\frac{1}{5} \int_{-5}^0 I(u) du \right), \text{ where } c = 0.8 \quad (3.22)$$

The intuition behind this barrier is that industry distress should depend on the average level of industry index. We then use the 5-year average of industry index as a benchmark. Distress then should mean that the industry index falls below a fraction of the past industry index. In this case we use a

constant $c = 0.8$. The value of c need not be fixed at 0.8. However, according to practitioners' views (Wall Street News, Yahoo Finance, etc.), the fall of stock by more than 20% signifies distress. Since most investors also listen to practitioners' views, we take $c = 0.8$ as the cutoff level.

Having defined industry distress barrier and the probability of distress given default, we are ready to price CDS. Before we do so, there is an issue about time-consistency of the model. In this simple model, we let $D(0)$ be the industry barrier, which is known at time 0. An investor will take $D(0)$ and use it to price CDS from year 0 to 5, forgetting all the past histories of the industry index. This assumption may be unrealistic. For example, the probability of distress looking from year 1 to year 5 may depend on the history of industry index from year -4 to year 0, and the dynamics of industry index from year 0 to year 1 which is still unknown at year 0. This is different from taking $D(0)$ as a constant and calculate the probability of distress from year 1 to year 5. With this concern, we also propose a different time-consistent model.

Time-Consistent Model

We modify the industry barrier dynamics to be a rolling average of the previous 5-year industry index. We first define the quantity

$$h(t) = \frac{1}{5} \int_{t-5}^t \log(I(u)) du \quad (3.23)$$

This is the average of industry index (in log scale) over the period of 5 years. We will specify industry distress as when

$$\log(I(t)) < b * h(t) \quad (3.24)$$

and b will be left to be determined later in the empirical section.

We focus first on the quantity $h(t)$. With this specification, $h(t)$ may depend on both the past information and the future randomness. For example

$$h(0) = \frac{1}{5} \int_{-5}^0 \log(I(u)) du$$

This will depend only on the past information. In other words, the probability of distress at time 0 is known by just comparing $\log(I(0))$ with $b * h(0)$

As the time moves forward, $h(t)$ will have randomness due to the randomness of the industry dynamics. For example

$$h(1) = \frac{1}{5} \int_{-4}^1 \log(I(u)) du = \frac{1}{5} \int_{-4}^0 \log(I(u)) du + \frac{1}{5} \int_0^1 \log(I(u)) du$$

The first quantity on the right side is known at time 0 and there is no randomness there. The second quantity will depend on the dynamics of $I(t)$ and is the source of randomness. We will call the first quantity $k(t)$ where

$$k(t) = \frac{1}{5} \int_{t-5}^0 \log(I(u)) du \quad (3.25)$$

Thus, we can write (3.23) as

$$h(t) = k(t) + \frac{1}{5} \int_0^t \log(I(u)) du \quad (3.26)$$

With the setup in (3.24) and (3.26), we can derive the probability of industry distress given the firm's default as follows:

$$P\{Distress(t)|Default(t)\} = N\left(\frac{\mu(t)}{\sqrt{t(1-|\rho|) + \frac{b^2 t^3}{75}}}\right) \quad (3.27)$$

where

$$\mu(t) = \begin{cases} \frac{bk(t) + (\frac{bt}{5} - 1)\log(I(0)) + (\frac{bt^2}{10} - t)(r - \frac{\sigma_I^2}{2})}{\sigma_I} + \sqrt{\rho}m & \text{if } \rho \geq 0 \\ \frac{bk(t) + (\frac{bt}{5} - 1)\log(I(0)) + (\frac{bt^2}{10} - t)(r - \frac{\sigma_I^2}{2})}{\sigma_I} - \sqrt{|\rho|}m & \text{if } \rho < 0 \end{cases}$$

Proof: See Appendix A.6

For shorthand, we call the above function p , i.e.,

$$P\{Distress(t)|Default(t)\} = p(m(0), k(t), I(0), \rho, t) \quad (3.28)$$

Note that we still need to determine the value of b for empirical tests. Obviously b must be less than 1, but the precise value of b will be determined later in the empirical section.

3.7.2 CDS Pricing with Stochastic Recovery Model

We have derived the expected recovery rates in the previous section. In this section we incorporate the result into the CDS pricing model.

Similar to the model in (3.2) by Hull et al. (2009), we let the recovery rate (R) be stochastic and get the model for CDS spreads as

$$S = \frac{\int_0^T (1 - E^Q[R(t)|Default(t)])q(u)v(u)du}{\int_0^T q(u)[h(u) + e(u)]du + (1 - \int_0^T q(u)du)h(T)} \quad (3.29)$$

We specify $E^Q[R(t)|Default(t)]$ in (3.15) and then $P\{Distress(t)|Default(t)\}$ in (3.21) and (3.28). Plugging in the results from the derivation, we get the following

1. The simple model:

$$\begin{aligned}
S &= \frac{\int_0^T (1 - a_1 + a_2 p(m(0), D(0), I(0), \rho, u)) q(u) v(u) du}{\int_0^T q(u) [h(u) + e(u)] du + (1 - \int_0^T q(u) du) h(T)} \\
&= (1 - a_1) \left(\frac{\int_0^T q(u) v(u) du}{\int_0^T q(u) [h(u) + e(u)] du + (1 - \int_0^T q(u) du) h(T)} \right) \\
&\quad + a_2 \left(\frac{\int_0^T p(m(0), D(0), I(0), \rho, u) q(u) v(u) du}{\int_0^T q(u) [h(u) + e(u)] du + (1 - \int_0^T q(u) du) h(T)} \right) \quad (3.30)
\end{aligned}$$

2. The time-consistent model:

$$\begin{aligned}
S &= \frac{\int_0^T (1 - a_1 + a_2 p(m(0), k(u), I(0), \rho, u)) q(u) v(u) du}{\int_0^T q(u) [h(u) + e(u)] du + (1 - \int_0^T q(u) du) h(T)} \\
&= (1 - a_1) \left(\frac{\int_0^T q(u) v(u) du}{\int_0^T q(u) [h(u) + e(u)] du + (1 - \int_0^T q(u) du) h(T)} \right) \\
&\quad + a_2 \left(\frac{\int_0^T p(m(0), k(u), I(0), \rho, u) q(u) v(u) du}{\int_0^T q(u) [h(u) + e(u)] du + (1 - \int_0^T q(u) du) h(T)} \right) \quad (3.31)
\end{aligned}$$

The equations (3.30) and (3.31) are similar except for the probability of distress given default on the RHS. The coefficient of interest here is a_2 which indicates how much the recovery rates fall if the firm defaults during the

period of industry distress. With our model specification, all parameters after the integral sign are observable. The only unknowns are a_1 and a_2 , with particular attention to a_2 . We proceed to estimate a_2 in the empirical section.

3.7.3 Empirical Analysis

We use the same CDS data as in the previous empirical part. The value of m is calculated from the DD_{BS} explained before. The only extra data required for this part are the industry index and the correlation between the industry index and the firm's value. For industry index, we use the data from Kenneth French's website¹. We use the monthly data for 10 industry portfolios. For each industry, we start with 100 as the industry index in 1994. Then we calculate the next industry index from the historical monthly returns. This way, we have historical industry index for 10 years before 2004, which is the beginning period of our CDS data. Note that we do not require the absolute size of the industry. We only need the relative size of the industry for each period so we can determine whether the industry is in distress with respect to its own past relative size.

To calculate the correlation between the industry index and the firm's value, we use the data from the same source. For industry returns, we use the data from Kenneth French's website for 10 industry portfolios. For firm's returns, we use the data from the CRSP database. We use 1-year past correlation of returns as an input to the model.

From (3.30) and (3.31), we can find the value after the integral sign by

¹http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

numerical integration. Since the model is exact, in the empirical analysis we write the model as

$$S = (1 - a_1)X_1 + a_2X_2 + \epsilon \quad (3.32)$$

where X_1 is the first bracket and X_2 is the second bracket of the RHS of (3.30) and (3.31), and ϵ is white noise. With this specification, we can simply estimate a_2 by an OLS regression. We select the industry distress ratio $c = 0.8$ for the simple model and $b = 0.94$ for the time-consistent model. The choice of $c = 0.8$ has been explained in the previous section. The choice of $b = 0.94$ is to be consistent with $c = 0.8$ in the linear scale. The regression result is shown in Table 3.11

The results show that the coefficient a_2 is highly significant. This means that the expected risk-neutral recovery rates are lower during the industry distress period. Obviously the magnitude of a_2 depends heavily on the distress barrier, i.e., the chosen value of c and b . If we choose $c = 0.8$ and $b = 0.94$, then the recovery rates are lower by about 20% if the firm defaults during the period of industry distress. Since in the regression we do not separate by industry, this number is an average of all industries. Note that this value is the expected decrease of recovery rates in the risk-neutral measure. It may not correspond exactly to the realized decrease of recovery rates in the physical measure, but we can still learn useful information about the variation of recovery rates across business cycles. Overall we conclude that our theoretical model complies with the data with reasonable parameters. The risk-neutral recovery rates are indeed stochastic and can be captured by this simple model.

3.8 Conclusion and Future Work

The recovery rates have long been treated as a constant in both academic and practitioner community. The literature on credit derivative pricing has focused mostly on estimating the probability of default. Recently it has been found that the realized recovery rates are stochastic and highly correlated with the industry condition when the firm defaults. We found that the market has already incorporated stochastic recovery rates into credit derivative pricing ex ante, in the risk-neutral measure. We derive a model to capture this finding which can be used to learn about stochastic recovery rates across business cycles.

Although it is hard to separate the effect of the risk-neutral probability of default and risk-neutral recovery rates, we have identified the effect of the risk-neutral recovery factor on CDS spreads. The market adjusts for the recovery rates ex ante to price credit derivatives. The ex ante recovery rates are consistent with the realized recovery rates found in the literature. In particular, recovery rates depend on industry characteristics and industry distress, both ex ante (in the risk-neutral measure) and ex post (in the physical measure).

There can be a few other explanations but we can rule them out. The first is that risk premium is time-varying and thus we may have not controlled for the risk-neutral default probabilities properly. In the robustness check section we address this concern and show that even if we let the default probabilities vary with time, the main result still goes through. The second is that illiquidity may be the true cause of industry characteristics and industry distress factors. We have included illiquidity in the robustness check and the

result still holds. The next question is the significance of the lag of industry distress factor. A possible explanation may be that the credit market is illiquid, so new information takes time to be incorporated into the price. The other explanation may be that price discovery occurs in the equity market long before the credit market. Thus using the industry distress information from the equity market to explain CDS spreads naturally exhibits the lag effect.

We propose a simple theoretical model to capture the stochastic recovery rates. The model links the unobserved risk-neutral recovery rates to the observable industry index. The expected risk-neutral recovery rate will be lower if the industry is in distress when the firm defaults. The model is a structural model with the first passage time as a default - the firm defaults the first time that its value crosses the debt barrier. The industry index also follows a structural model - the industry is in distress if the industry index is below the industry barrier. With a reasonable set of parameters, the model is consistent with the data and the expected risk-neutral recovery rate is lower by about 20% if the firm defaults when the industry is in distress. However, it is the structure of the model that we emphasize here, not the number 20%. We can use this model to learn about stochastic recovery rates in different periods and may get different numbers from 20%.

The stochastic recovery rates are indeed priced in the CDS spreads. Our theoretical model captures the empirical findings and can be used to price credit derivatives. We can also use the model to extract expected recovery rates across business cycles and use this information for risk management. The next issue may be to use this stochastic recovery model to price Collat-

eralized Debt Obligations (CDOs) and study the effect of stochastic recovery rates on the implied correlation smile.

3.9 Figures and Tables

Table 3.1: Summary Statistics of CDS data

The table shows the summary statistics of CDS data used in the empirical section. *CDS Spread* shows the spreads in basis points. The remaining information in the table comes from the CRSP database. *Market Cap* shows the market capitalization of firms in millions. *Stock Monthly Volatility* shows monthly volatility of the stock. *Beta*(β) shows the beta coefficient when regressing the stock returns with market returns. *SMB* and *HML* shows the regression coefficients when regressing the stock returns with the *SMB* and *HML* factors

	CDS Spread	Market Cap (M)	Stock Monthly Volatility	β	SMB	HML
mean	79.27	27,496	0.068	0.94	0.12	0.31
median	38.60	13,524	0.062	0.85	0	0.19
min	1.5	176	0.031	-0.19	-3.05	-1.34
5% quantile	11.7	1,923	0.036	0.16	-0.87	-0.81
25% quantile	23.7	6,404	0.049	0.63	-0.33	-0.19
50% quantile	38.6	13,524	0.062	0.85	0	0.19
75% quantile	70.5	27,955	0.078	1.18	0.60	0.76
95% quantile	316.1	99,429	0.120	2.09	1.41	1.67
max	2666.5	513,362	0.229	2.77	2.44	2.79

Table 3.2: Industry Classification

This table shows the percentage of firms in the data set classified into each industry, according to the Siccodes on Kenneth French's website.

Number	Industry	Percent
1	Consumer Nondurables	7.75
2	Consumer Durables	2.87
3	Manufacturing	17.65
4	Energy	5.41
5	Hi-tech	5.69
6	Telecom	3.10
7	Wholesale, Retail	12.48
8	Healthcare	7.20
9	Utilities	8.76
10	Other: Mines, Trans, Const, Finance, etc	29.10

Table 3.3: Distress cutoff returns for each industry

The table shows the monthly returns below which the industry is in distress. The cutoff level is the 10th percentile of historical monthly industry returns from 1994 to 2003. The industry returns data are from Kenneth French's website.

Number	Fama-French Industry	Distress Cutoff
1	Consumer Nondurables	-4.1%
2	Consumer Durables	-6.7%
3	Manufacturing	-4.6%
4	Energy	-4.8%
5	Hi-tech	-11.8%
6	Telecom	-8.4%
7	Wholesale, Retail	-5.0%
8	Healthcare	-5.8%
9	Utilities	-5.3%
10	Other: Mines, Trans, Const, Finance, etc	-5.0%

Table 3.4: OLS regression for recovery rates effect in CDS
The dependent variable is $\log(\text{CDS spreads})$. $I1 - I9$ are industry dummy variables for industry 1 to 9. $Distress$ and $lag(Distress)$ are dummy variables for industry distress condition as defined in Table 3.3. PD and PD from ratings are probability of default calculated from the PD_{BS} and S&P Ratings. Interest rates of Treasury bills for 3 months and 5 years are also included as control variables. Standard errors are adjusted for 10 clusters in industry and shown in parentheses. (***)significant at 1% level, ** significant at 5% level, * significant at 10% level)

Explanatory Variables	With Industry Condition	Without Industry Condition
Intercept	5.33*** (0.17)	
Distress	0.40** (0.15)	
lag(Distress)	0.39*** (0.11)	
I1 (Consumer Nondurables)	-0.17*** (0.014)	
I2 (Consumer Durables)	0.14*** (0.017)	
I3 (Manufacturing)	-0.17*** (0.014)	
I4 (Energy)	-0.30*** (0.008)	

Continued on next page

Table 3.4 – *Continued from previous page*

Explanatory Variables	With Industry Condition	Without Industry Condition
I5 (Hi-tech)	−0.11*** (0.032)	
I6 (Telecom)	0.35*** (0.039)	
I7 (Wholesale, Retail)	−0.15*** (0.011)	
I8 (Healthcare)	0.04 (0.04)	
I9 (Utilities)	−0.22*** (0.024)	
log(PD)	0.0015*** (0.0003)	0.0016*** (0.0003)
log(PD from Ratings)	0.43*** (0.034)	0.42*** (0.033)
3-month Int. Rates	0.02 (0.02)	0.056 (0.043)
5-year Int. Rates	−0.35*** (0.06)	−0.44*** (0.11)
Obs	13477	13477
R^2	0.6386	0.6047

Table 3.5: OLS regression for recovery rates effect in CDS with firm and time fixed effects

The dependent variable is $\log(\text{CDS spreads})$. $I1 - I9$ are industry dummy variables for industry 1 to 9. $Distress$ and $\text{lag}(Distress)$ are dummy variables for industry distress condition as defined in Table 3.3. PD and $PD \text{ from ratings}$ are probability of default calculated from the PD_{BS} and S&P Ratings. Interest rates of Treasury bills for 3 months and 5 years are also included as control variables. Standard errors are adjusted for 10 clusters in industry and shown in parentheses. (***)significant at 1% level, ** significant at 5% level, * significant at 10% level)

Fixed Effect	None	Firm and Year	Firm and Month
Intercept	5.33*** (0.17)	3.94*** (0.086)	5.40*** (0.26)
Distress	0.40** (0.15)	0.33** (0.12)	0.16** (0.06)
lag(Distress)	0.39*** (0.11)	0.32*** (0.07)	0.16** (0.07)
I1	-0.17*** (0.014)	0.46*** (0.12)	0.41*** (0.13)
I2	0.14*** (0.017)	-0.008 (0.16)	0.04 (0.17)
I3	-0.17*** (0.014)	-0.11* (0.05)	-0.09* (0.04)
I4	-0.30***	-0.49**	-0.48**

Continued on next page

Table 3.5 – *Continued from previous page*

Fixed Effect	None	Firm and Year	Firm and Month
	(0.008)	(0.18)	(0.20)
I5	−0.11***	0.16	0.13
	(0.032)	(0.13)	(0.15)
I6	0.35***	0.36***	0.32**
	(0.039)	(0.11)	(0.12)
I7	−0.15***	−0.31***	−0.30***
	(0.011)	(0.07)	(0.07)
I8	0.04	0.29**	0.26**
	(0.04)	(0.10)	(0.11)
I9	−0.22***	−0.10	−0.08
	(0.024)	(0.10)	(0.10)
log(PD)	0.0015***	0.0012***	0.0011***
	(0.0003)	(0.00013)	(0.00012)
log(PD from Ratings)	0.43***	0.35***	0.35***
	(0.034)	(0.065)	(0.068)
3-month Int. Rates	0.02	−0.18***	0.25***
	(0.02)	(0.03)	(0.07)
5-year Int. Rates	−0.35***	0.006	−0.62***
	(0.06)	(0.015)	(0.12)
Obs	13477	13477	13477
R^2	0.6386	0.8294	0.8466

Table 3.6: OLS regression for recovery rates effect in CDS, with control variables, but without fixed effects

The independent variables are the same as Table 3.4.

Extra control variables are as follows: *Index Returns* are S&P500 index returns. *Industry Returns* are the monthly returns for each industry. *log(Bid-Ask Spreads)* is the log of bid-ask spreads of CDS quotes. Standard errors are adjusted for 10 clusters in industry and shown in parentheses. (***) significant at 1% level, ** significant at 5% level, * significant at 10% level)

Explanatory Variables	Model			
	Original	with Index Returns	with Industry Returns	with Illiquidity
Intercept	5.33*** (0.17)	5.32*** (0.18)	5.35*** (0.17)	3.61*** (0.09)
Distress	0.40** (0.15)	0.31* (0.15)	0.28** (0.13)	0.20*** (0.020)
				3.62*** (0.09)
				0.13 (0.08)

Continued on next page

Table 3.6 – Continued from previous page

Explanatory Variables	Original	with Index Returns	with Industry Returns	with Illiquidity	All
lag(Distress)	0.39*** (0.11)	0.38*** (0.11)	0.37*** (0.10)	0.19*** (0.020)	0.19*** (0.05)
I1	-0.17*** (0.014)	-0.17*** (0.014)	-0.16*** (0.015)	-0.14*** (0.011)	-0.14*** (0.011)
I2	0.14*** (0.017)	0.14*** (0.017)	0.13*** (0.017)	0.014 (0.018)	0.015 (0.018)
I3	-0.17*** (0.014)	-0.17*** (0.014)	-0.16*** (0.016)	-0.19*** (0.009)	-0.19*** (0.011)
I4	-0.30*** (0.008)	-0.30*** (0.008)	-0.27*** (0.01)	-0.23*** (0.01)	-0.23*** (0.007)
I5	-0.11*** (0.032)	-0.11*** (0.032)	-0.11*** (0.031)	-0.14*** (0.022)	-0.14*** (0.022)
I6	0.35*** (0.039)	0.34*** (0.039)	0.34*** (0.039)	0.19*** (0.023)	0.18*** (0.023)
I7	-0.15*** (0.039)	-0.15*** (0.039)	-0.15*** (0.039)	-0.08*** (0.023)	-0.08*** (0.023)

Continued on next page

Table 3.6 – Continued from previous page

Explanatory Variables	Original	with Index Returns	with Industry Returns	with Illiquidity	All
I8	(0.011) 0.04 (0.04)	(0.011) 0.04 (0.04)	(0.011) 0.04 (0.04)	(0.012) −0.096** (0.033)	(0.013) −0.099** (0.033)
I9	−0.22*** (0.024)	−0.22*** (0.02)	−0.20*** (0.026)	−0.23*** (0.018)	−0.23*** (0.02)
log(PD)	0.0015*** (0.0003)	0.0015*** (0.0003)	0.0014*** (0.0003)	0.0014*** (0.0002)	0.0014*** (0.0002)
log(PD from Ratings)	0.43*** (0.034)	0.43*** (0.034)	0.43*** (0.034)	0.28*** (0.031)	0.28*** (0.032)
3-month Int Rates	0.02 (0.02)	0.02 (0.03)	0.02 (0.03)	0.028 (0.016)	0.029 (0.016)
5-year Int Rates	−0.35*** (0.06)	−0.35*** (0.06)	−0.36*** (0.06)	−0.21*** (0.024)	−0.21*** (0.024)
Index Returns		−2.36*** (0.24)			−1.19** (0.417)

Continued on next page

Table 3.6 – Continued from previous page

Explanatory Variables	Original	with Index Returns	with Industry Returns	with Illiquidity	All
Industry Returns			−0.015*** (0.004)		−0.003 (0.0035)
log(Bid-Ask Spreads)				0.69*** (0.037)	0.68*** (0.038)
Obs	13477	13477	13477	13463	13463
R^2	0.6386	0.6418	0.6407	0.7559	0.7571

Table 3.7: OLS regression for recovery rates effect in CDS, with control variables, with fixed effects

The independent variables are the same as Table 3.4. Extra control variables are as follows: *Index Returns* are S&P500 index returns. *Industry Returns* are the monthly returns for each industry. *log(Bid-Ask Spreads)* is the log of bid-ask spreads of CDS quotes. Standard errors are adjusted for 10 clusters in industry and shown in parentheses. (***)significant at 1% level, ** significant at 5% level, * significant at 10% level)

	Fixed Effect			
Explanatory Variables	None	Firm Only	Firm and Year	Firm and Month
Intercept	3.62*** (0.09)	3.71*** (0.18)	3.00*** (0.18)	4.00*** (0.12)
Distress	0.13 (0.08)	0.15* (0.08)	0.14* (0.08)	0.09** (0.035)
lag(Distress)	0.19*** (0.05)	0.22*** (0.045)	0.20*** (0.03)	0.09** (0.037)
I1	-0.14*** (0.011)	0.39*** (0.05)	0.43*** (0.06)	0.39*** (0.07)
I2	0.015 (0.018)	0.004 (0.07)	0.06 (0.11)	0.09 (0.12)
I3	-0.19***	-0.04	-0.03	-0.04

Continued on next page

Table 3.7 – *Continued from previous page*

Explanatory Variables	None	Firm Only	Firm and Year	Firm and Month
	(0.011)	(0.06)	(0.05)	(0.04)
I4	−0.23***	−0.19**	−0.32**	−0.33**
	(0.007)	(0.07)	(0.11)	(0.13)
I5	−0.14***	0.08	0.07	0.04
	(0.022)	(0.06)	(0.07)	(0.09)
I6	0.18***	0.28**	0.23**	0.22**
	(0.023)	(0.10)	(0.09)	(0.10)
I7	−0.08***	−0.20***	−0.16***	−0.15**
	(0.013)	(0.06)	(0.05)	(0.05)
I8	−0.099***	0.33***	0.27***	0.24**
	(0.033)	(0.09)	(0.07)	(0.08)
I9	−0.23***	−0.02	−0.07	−0.05
	(0.02)	(0.06)	(0.07)	(0.07)
log(PD)	0.0014***	0.0010***	0.0010***	0.0010***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
log(PD from Ratings)	0.28***	0.28***	0.29***	0.29***
	(0.032)	(0.06)	(0.06)	(0.06)
3-month Int Rates	0.029	0.03*	−0.1***	0.20***
	(0.016)	(0.016)	(0.02)	(0.03)
5-year Int Rates	−0.21***	−0.25***	0.01***	−0.44***
	(0.024)	(0.024)	(0.01)	(0.044)
Index Returns	−1.19**	−1.43**	−1.13***	−1.14***

Continued on next page

Table 3.7 – *Continued from previous page*

Explanatory Variables	None	Firm Only	Firm and Year	Firm and Month
Industry Returns	(0.417)	(0.5)	(0.23)	(0.27)
	−0.003	−0.004	−0.0017	−0.0017
	(0.0035)	(0.004)	(0.0017)	(0.002)
log(Bid-Ask Spreads)	0.68***	0.53***	0.50***	0.47***
	(0.038)	(0.07)	(0.05)	(0.05)
Obs	13463	13463	13463	13463
R^2	0.7571	0.8579	0.8684	0.8793

Table 3.8: OLS regression for recovery rates effect in CDS with control variables and fixed effects, before and during the crisis

The independent variables are the same as in Table 3.6. Before-Crisis is the period 2004-2006. During-Crisis is the period 2007-2008. We drop firms with the market cap more than 5% of the industry. Standard errors are adjusted for 10 clusters in industry and shown in parentheses. (***)significant at 1% level, ** significant at 5% level, * significant at 10% level)

Explanatory Variables	Period	
	Before Crisis	During Crisis
Distress	0.001 (0.05)	0.16*** (0.04)
lag(Distress)	0.12** (0.05)	0.15*** (0.03)
log(PD)	0.0006*** (0.0001)	0.0006** (0.0002)
log(PD from Ratings)	0.37*** (0.08)	0.12*** (0.033)
Obs	8439	3762
R^2	0.8816	0.9263

Table 3.9: OLS regression for recovery rates effect in CDS with control variables and fixed effects, with time-varying risk-neutral probability of default

The independent variables are the same as in Table 3.6. We include the interaction terms between probability of default and time to control for time-varying risk-neutral probability of default. Standard errors are adjusted for 10 clusters in industry and shown in parentheses. (***) significant at 1% level, ** significant at 5% level, * significant at 10% level)

	Interaction Terms
Explanatory Variables	Yearly
Distress	0.12 (0.08)
lag(Distress)	0.22*** (0.05)
log(PD)	0.0015** (0.0005)
log(PD)*Year2005	-0.0009** (0.0004)
log(PD)*Year2006	-0.0009** (0.0004)
log(PD)*Year2007	-0.0004

Continued on next page

Table 3.9 – *Continued from previous page*

Explanatory Variables	Yearly
	(0.0007)
$\log(\text{PD}) * \text{Year2008}$	0.0006
	(0.0010)
$\log(\text{PD from Ratings})$	0.30***
	(0.05)
$\log(\text{PD from Ratings}) * \text{Yr2005}$	0.008
	(0.02)
$\log(\text{PD from Ratings}) * \text{Yr2006}$	0.07***
	(0.02)
$\log(\text{PD from Ratings}) * \text{Yr2007}$	−0.009
	(0.02)
$\log(\text{PD from Ratings}) * \text{Yr2008}$	−0.15***
	(0.04)
Obs	13463
R^2	0.8728

Table 3.10: OLS regression for recovery rates effect in CDS with control variables and fixed effects, using implied volatilities.

The independent variables are the same as Table 3.4. Extra control variables are as follows: *Index Returns* are S&P500 index returns. *Industry Returns* are the monthly returns for each industry. *log(Bid-Ask Spreads)* is the log of bid-ask spreads of CDS quotes. Standard errors are adjusted for 10 clusters in industry and shown in parentheses. (***)significant at 1% level, ** significant at 5% level, * significant at 10% level)

Explanatory Variables	Fixed Effect			
	None	Firm Only	Firm and Year	Firm and Month
Distress	0.11 (0.07)	0.14* (0.07)	0.13 (0.08)	0.10** (0.036)
lag(Distress)	0.17*** (0.04)	0.20*** (0.03)	0.20*** (0.03)	0.11** (0.038)
log(PD from Implied Volatility)	0.0028*** (0.0009)	0.003*** (0.0005)	0.003*** (0.0005)	0.003*** (0.0004)
log(PD from Ratings)	0.27*** (0.03)	0.26*** (0.06)	0.26*** (0.06)	0.27*** (0.06)
Obs	13309	13309	13309	13309
R^2	0.7600	0.8640	0.8715	0.8802

Table 3.11: OLS regression for the CDS pricing model with stochastic recovery rates - test for the industry distress effect

The independent variables X_1 and X_2 are defined as in (3.32). The corresponding coefficients are $(1 - a_1)$ and a_2 . The industry distress effect on recovery rates is shown through the coefficient a_2 of the variable X_2 . The first column is the simple model as in (3.30). The second column is the time-consistent model as in (3.31). Standard errors are in parentheses. (***) significant at 1% level, ** significant at 5% level, * significant at 10% level)

Explanatory Variables	Model	
	Simple	Time-Consistent
Constant	66.45*** (0.84)	66.41*** (0.84)
X_1	0.14*** (0.0019)	0.12*** (0.0024)
X_2	0.24*** (0.0138)	0.20*** (0.0105)
Obs	14129	14135
R^2	0.32	0.33

Appendix A

Proofs

A.1 First Passage Time Density

Assume

$$\frac{dV}{V} = rdu + \sigma dW^Q$$

Then

$$\frac{V(u)}{V(0)} = e^{(r - \frac{\sigma^2}{2})u + \sigma W(u)}$$

where $W(u) \sim N(0, u)$

Thus,

$$W(u) = \frac{\log(\frac{V(u)}{V(0)}) - (r - \frac{\sigma^2}{2})u}{\sigma} \quad (\text{A.1})$$

Let $B(u)$ be the default barrier at time u . By our assumption, $B(u) = B(0)e^{(r - \frac{\sigma^2}{2})u}$

The default time is $\min \{u : V(u) = B(u)\}$. Plug $B(u)$ in $V(u)$ in (A.1), we get the corresponding

$$W(u) = \frac{\log(\frac{B(0)e^{(r - \frac{\sigma^2}{2})u}}{V(0)}) - (r - \frac{\sigma^2}{2})u}{\sigma} = \frac{\log(\frac{B(0)}{V(0)})}{\sigma} = \tilde{m} \text{ (constant)}$$

Note that $\min \{u : V(u) = B(u)\} = \min \{u : W(u) = \tilde{m}\}$. We can thus calculate the default density by using the first passage time of $W(u)$ to \tilde{m} . Since Brownian motion is symmetric, the passage time to level \tilde{m} is equivalent to the passage time to level $|\tilde{m}|$. For convenience, we consider $m = |\tilde{m}| = \frac{\log(\frac{V(0)}{B(0)})}{\sigma}$ instead.

We define the first passage time to level m :

$$\tau_m = \min \{u : X(u) = m\}$$

where $m = \frac{\log(\frac{V(0)}{B(0)})}{\sigma}$ and $W(0) = 0$

Since this is the first passage time of Brownian motion without drift across a constant barrier m , we use the result from Shreve (2004), theorem 3.7.1. τ_m has cumulative distribution function

$$P \{ \tau_m \leq u \} = 2\Phi\left(\frac{-m}{\sqrt{u}}\right)$$

and density

$$q(m, u) = \frac{d}{du} P \{ \tau_m \leq u \} = \frac{m}{u\sqrt{2\pi u}} e^{\frac{-m^2}{2u}}$$

A.2 Dynamic of Distance-to-Default

The dynamics of V follows a standard geometric Brownian motion:

$$\frac{dV}{V} = rdt + \sigma dW^Q$$

Consider $m = \frac{\log(\frac{V}{B})}{\sigma}$, thus $\frac{\partial m}{\partial V} = \frac{1}{\sigma V}$, $\frac{\partial m}{\partial B} = -\frac{1}{\sigma B}$ and $\frac{\partial^2 m}{\partial V^2} = -\frac{1}{\sigma V^2}$. We then write equation (1.4) as:

$$\begin{aligned} dm &= \frac{1}{\sigma V}(rVdt + \sigma VdW^Q) + \frac{1}{2}\left(-\frac{1}{\sigma V^2}\right)(\sigma V)^2dt + \left(-\frac{1}{\sigma B}\right)\left(B\left(r - \frac{\sigma^2}{2}\right)\right)dt \\ &= \frac{r}{\sigma}dt + dW^Q - \frac{\sigma}{2}dt + \left(-\frac{r}{\sigma} + \frac{\sigma}{2}\right)dt \\ &= dW^Q \end{aligned}$$

Note that the volatiitiy (σ) and drift (μ) terms all cancel out on the second line.

A.3 Analytical Approach

The analytical formula for $\frac{\partial S}{\partial m}$ and $\frac{\partial^2 S}{\partial m^2}$ can be derived as follows. Take $S(m)$ from (1.2), we can write:

$$S(m) = \frac{(1 - \hat{R})(1 + a^*) \int_0^T q(m, r)v(r)dr}{\int_0^T q(m, r)[h(r) + e(r)]dr + (1 - \int_0^T q(m, r)dr)h(T)} = \frac{f(m)}{g(m)} \quad (\text{A.2})$$

Then,

$$\frac{\partial S}{\partial m} = \frac{g \frac{\partial f}{\partial m} - f \frac{\partial g}{\partial m}}{g^2} \quad (\text{A.3})$$

where

$$\frac{\partial f}{\partial m} = (1 - \hat{R})(1 + a^*) \int_0^T \frac{\partial q(m, r)}{\partial m} v(r)dr$$

and

$$\frac{\partial g}{\partial m} = \int_0^T \frac{\partial q(m, r)}{\partial m} [h(r) + e(r)]dr - \left(\int_0^T \frac{\partial q(m, r)}{\partial m} dr \right) h(T)$$

Now consider $q(m, r) = \frac{m}{r\sqrt{2\pi r}} e^{-\frac{m^2}{2r}}$, we get

$$\frac{\partial q(m, r)}{\partial m} = \frac{1}{r\sqrt{2\pi r}} e^{-\frac{m^2}{2r}} + \frac{m}{r\sqrt{2\pi r}} \frac{-m}{r} e^{-\frac{m^2}{2r}} = \left(\frac{r - m^2}{r^2\sqrt{2\pi r}} \right) e^{-\frac{m^2}{2r}} \quad (\text{A.4})$$

We plug the value from (A.4) back to $\frac{\partial f}{\partial m}$ and $\frac{\partial g}{\partial m}$ and get

$$\frac{\partial f}{\partial m} = (1 - \hat{R})(1 + a^*) \int_0^T \left(\frac{r - m^2}{r^2 \sqrt{2\pi r}} \right) e^{\frac{-m^2}{2r}} v(r) dr \quad (\text{A.5})$$

$$\frac{\partial g}{\partial m} = \int_0^T \left(\frac{r - m^2}{r^2 \sqrt{2\pi r}} \right) e^{\frac{-m^2}{2r}} [h(r) + e(r)] dr - \left(\int_0^T \left(\frac{r - m^2}{r^2 \sqrt{2\pi r}} \right) e^{\frac{-m^2}{2r}} dr \right) h(T) \quad (\text{A.6})$$

We then plug (A.5) and (A.6) back into (A.3) and get expression for $\frac{\partial S}{\partial m}(m)$. Note that this only depends on m which only depends on the ratio $V(t)/B(t)$ at the time of interest. One can also find $\frac{\partial^2 S}{\partial m^2}$ by applying quotient rule again to $\frac{\partial S}{\partial m}$ and get

$$\frac{\partial^2 S}{\partial m^2} = \frac{g \frac{\partial^2 f}{\partial m^2} - f \frac{\partial^2 g}{\partial m^2} - 2 \frac{\partial f}{\partial m} \frac{\partial g}{\partial m}}{g^2} + \frac{2f \left(\frac{\partial g}{\partial m} \right)^2}{g^3}$$

A.4 First Passage Time Density of Mean-Reverting Process

This part only replicates the result from Alili et al. (2005). In essence, we do not prove a new result here, but the transformation of variables is needed to be able to apply the formula correctly.

The associated Ornstein-Uhlenbeck process $(U_t)_{t \geq 0}$ with parameter λ is defined to be the unique solution to the stochastic differential solution

$$dU_t = -\lambda U_t dt + dB_t \quad (\text{A.7})$$

and $U_0 = x$. For a fixed real a , introduce the stopping time

$$\sigma_a = \inf \{s > 0 : U_s = a\} \quad (\text{A.8})$$

Its law is absolutely continuous with respect to the Lebesgue measure. We set

$$\mathbf{P}_x^{(\lambda)}(\sigma_a \in dt) = p_{x \rightarrow a}^{(\lambda)}(t)dt \quad (\text{A.9})$$

Theorem. For any $x < a$, we have the series expansion

$$p_{x \rightarrow a}^{(\lambda)}(t) = -\lambda e^{\lambda(x^2 - a^2)/2} \sum_{j=1}^{\infty} \frac{D_{j, -a\sqrt{2\lambda}}(-x\sqrt{2\lambda})}{D'_{j, -a\sqrt{2\lambda}}(-a\sqrt{2\lambda})} \exp(-\lambda v_{j, -a\sqrt{2\lambda}} t) \quad (\text{A.10})$$

where $D_v(\cdot)$ is the parabolic cylinder function with index v , and $D'_{v_{j,b}}(b) = \left. \frac{\partial D_v(b)}{\partial v} \right|_{v=v_{j,b}}$.

For $v \rightarrow \infty$, we have the asymptotic formula

$$D_v(z) \approx \sqrt{2}(v+1/2)^{v/2} e^{-(v+1/2)/2} \times \cos(z\sqrt{v+1/2} - \pi v/2)(1 + O(v^{-1/2}))$$

The following large- n asymptotics can be deduced

$$v_{n, -a\sqrt{2\lambda}} \approx 2n - 1 + 4\frac{\lambda a^2}{\pi^2} - 2\frac{\sqrt{\lambda}a}{\pi} \sqrt{4n - 1 + 4\frac{\lambda a^2}{\pi^2}}$$

and

$$\frac{D_{j, -a\sqrt{2\lambda}}(-x\sqrt{2\lambda})}{D'_{j, -a\sqrt{2\lambda}}(-a\sqrt{2\lambda})} \approx (-1)^n 2 \frac{\sqrt{2v_{n, -a\sqrt{2\lambda}} + 1}}{\pi \sqrt{2v_{n, -a\sqrt{2\lambda}} + 2a\sqrt{\pi}}} \times \cos\left(x\sqrt{\lambda(2v_{n, -a\sqrt{2\lambda}} + 1)} + \frac{\pi v_{n, -a\sqrt{2\lambda}}}{2}\right)$$

The formula above is formulated for the mean-reverting process with the target 0 and the initial position x . The barrier for the first passage time is $a > x$. To fit our process of m to the formula, we need to change the variable as follows:

$$U = m - \bar{m}^Q$$

Then

$$dU_t = -\lambda U_t dt + dB_t$$

as desired. The initial condition is $U_0 = x = m_0 - \bar{m}^Q$. The barrier is $a = 0 - \bar{m}^Q = -\bar{m}^Q$.

With this change of variable, we can then use the close form solution (A.10) to calculate the first passage time density and compute the CDS spread accordingly.

A.5 Probability of Distress – Simple Model

From the setup, we can write that

$$I(t) = I(0)e^{(r-\sigma_I^2/2)t+\sigma_I Z(t)} \quad (\text{A.11})$$

We define industry distress as $I(t) < D(0)$. From Appendix A, default occurs when $W(t) = -m$. Thus

$$\begin{aligned} P\{Distress(t)|Default(t)\} &= P\{I(t) < D(0)|W(t) = -m\} \\ &= P\left\{I(0)e^{(r-\sigma_I^2/2)t+\sigma_I Z(t)} < D(0)|W(t) = -m\right\} \\ &= P\left\{Z(t) < \frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sigma_I} \mid \right. \\ &\quad \left. W(t) = -m\right\} \end{aligned} \quad (\text{A.12})$$

Since $\text{corr}(dZ, dW) = \rho$, we let

$$Z(t) = \begin{cases} \sqrt{\rho}W(t) + \sqrt{1-\rho}X(t) & \text{if } \rho \geq 0 \\ -\sqrt{|\rho|}W(t) + \sqrt{1-|\rho|}X(t) & \text{if } \rho < 0 \end{cases}$$

where $X(t)$ is another Brownian motion independent of $Z(t)$ and $W(t)$.

We then plug the decomposition of $Z(t)$ into (A.12). We demonstrate the

case where $\rho \geq 0$. The case where $\rho < 0$ will be summarized at the end.

$$\begin{aligned}
P\{Distress(t)|Default(t)\} &= P\left\{\sqrt{\rho}W(t) + \sqrt{1-\rho}X(t) < \right. \\
&\quad \left. \frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sigma_I} | W(t) = -m \right\} \\
&= P\left\{ -\sqrt{\rho}m + \sqrt{1-\rho}X(t) < \right. \\
&\quad \left. \frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sigma_I} \right\} \\
&= P\left\{ X(t) < \frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sqrt{1-\rho}\sigma_I} \right. \\
&\quad \left. + \frac{\sqrt{\rho}m}{\sqrt{1-\rho}} \right\} \\
&= P\left\{ \frac{X(t)}{\sqrt{t}} < \frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sqrt{t}\sqrt{1-\rho}\sigma_I} \right. \\
&\quad \left. + \frac{\sqrt{\rho}m}{\sqrt{t}\sqrt{1-\rho}} \right\} \\
&= N\left(\frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sqrt{t}\sqrt{1-\rho}\sigma_I} \right. \\
&\quad \left. + \frac{\sqrt{\rho}m}{\sqrt{t}\sqrt{1-\rho}} \right) \tag{A.13}
\end{aligned}$$

If $\rho < 0$, then the derivation is similar and the final result is

$$P\{Distress(t)|Default(t)\} = N\left(\frac{\log(D(0)/I(0)) - (r - \sigma_I^2/2)t}{\sqrt{t}\sqrt{1-|\rho|}\sigma_I} - \frac{\sqrt{|\rho|m}}{\sqrt{t}\sqrt{1-|\rho|}}\right) \tag{A.14}$$

A.6 Probability of Distress – Time Consistent Model

First we note that $h(t)$ is a sum of normal random variables, and so is normally distributed. We need to find the mean and variance of $h(t)$

$$\begin{aligned}
 E[h(t)] &= E[k(t) + \frac{1}{5} \int_0^t \log(I(u)) du] \\
 &= k(t) + \frac{1}{5} E[\int_0^t \log(I(u)) du] \\
 &= k(t) + \frac{1}{5} \int_0^t E[\log(I(u))] du \\
 &= k(t) + \frac{1}{5} \int_0^t \log(I(0)) + (r - \frac{\sigma_I^2}{2}) u du \\
 &= k(t) + \frac{1}{5} \left(\log(I(0)) t + (r - \frac{\sigma_I^2}{2}) \frac{t^2}{2} \right)
 \end{aligned}$$

$$\begin{aligned}
Var[h(t)] &= Var[k(t) + \frac{1}{5} \int_0^t \log(I(u)) du] \\
&= \frac{1}{25} Var[\int_0^t \log(I(u)) du] \\
&= \frac{1}{25} Var[\int_0^t \log(I(0)) + (r - \frac{\sigma_I^2}{2})u + \sigma_I Z_u du] \\
&= \frac{1}{25} Var[\int_0^t \sigma_I Z_u du] \\
&= \frac{\sigma_I^2}{25} E \left[\int_0^t Z_u du \int_0^t Z_s ds \right] \\
&= \frac{\sigma_I^2}{25} E \left[\int_0^t \int_0^t Z_u Z_s dud s \right] \\
&= \frac{\sigma_I^2}{25} \int_0^t \int_0^t E[Z_u Z_s] dud s \\
&= \frac{\sigma_I^2}{25} \int_0^t \int_0^t \min(u, s) dud s \\
&= \frac{2\sigma_I^2}{25} \int_0^t \int_0^s u dud s \\
&= \frac{2\sigma_I^2}{25} \int_0^t \frac{s^2}{2} ds \\
&= \frac{\sigma_I^2}{25} \frac{t^3}{3}
\end{aligned}$$

Having calculated the mean and variance of $h(t)$, we are now ready to calculate the probability of distress given firm's default. First we signify $h(t)$

as a normal random variable

$$h(t) = k(t) + \frac{1}{5}(\log(I(0))t + (r - \frac{\sigma_I^2}{2})\frac{t^2}{2}) + \frac{\sigma_I}{5}B(t) \quad (\text{A.15})$$

where

$$B(t) \sim N(0, t^3/3)$$

Then

$$\begin{aligned} P\{Distress(t)|Default(t)\} &= P\{\log(I(t)) < bh(t)|W(t) = -m\} \\ &= P\left\{\log(I(0)) + (r - \frac{\sigma_I^2}{2})t + \sigma_I Z(t) < b\left(k(t) + \frac{1}{5}(\log(I(0))t + (r - \frac{\sigma_I^2}{2})\frac{t^2}{2}) + \frac{\sigma_I}{5}B(t)\right)\right. \\ &\quad \left.|W(t) = -m\right\} \\ &= P\left\{Z(t) < \frac{b\left(k(t) + \frac{1}{5}(\log(I(0))t + (r - \frac{\sigma_I^2}{2})\frac{t^2}{2})\right) - \log(I(0)) - (r - \frac{\sigma_I^2}{2})t}{\sigma_I} + \frac{b}{5}B(t)\right. \\ &\quad \left.|W(t) = -m\right\} \end{aligned}$$

Again, we decompose $Z(t)$ into the correlated and uncorrelated parts with the firm. Since $\text{corr}(dZ, dW) = \rho$, we let

$$Z(t) = \begin{cases} \sqrt{\rho}W(t) + \sqrt{1-\rho}X(t) & \text{if } \rho \geq 0 \\ -\sqrt{|\rho|}W(t) + \sqrt{1-|\rho|}X(t) & \text{if } \rho < 0 \end{cases}$$

where $X(t)$ is another Brownian motion independent of $Z(t)$ and $W(t)$.

We carry on the calculation for the case where $\rho \geq 0$. The case where $\rho < 0$ is similar and the result will be provided at the end. Substituting the

decomposition of $Z(t)$ into the previous equation, we get

$$\begin{aligned}
&= P \left\{ \sqrt{\rho}W(t) + \sqrt{1-\rho}X(t) < \frac{b \left(k(t) + \frac{1}{5}(\log(I(0)))t + (r - \frac{\sigma_I^2}{2})\frac{t^2}{2} \right)}{\sigma_I} \right. \\
&\quad \left. \frac{-\log(I(0)) - (r - \frac{\sigma_I^2}{2})t}{\sigma_I} + \frac{b}{5}B(t) | W(t) = -m \right\} \\
&= P \left\{ -\sqrt{\rho}m + \sqrt{1-\rho}X(t) < \frac{b \left(k(t) + \frac{1}{5}(\log(I(0)))t + (r - \frac{\sigma_I^2}{2})\frac{t^2}{2} \right)}{\sigma_I} \right. \\
&\quad \left. \frac{-\log(I(0)) - (r - \frac{\sigma_I^2}{2})t}{\sigma_I} + \frac{b}{5}B(t) \right\} \\
&= P \left\{ \sqrt{1-\rho}X(t) < \frac{\left(bk(t) + (\frac{bt}{5} - 1)\log(I(0)) + (\frac{bt^2}{10} - t)(r - \frac{\sigma_I^2}{2}) \right)}{\sigma_I} \right. \\
&\quad \left. + \sqrt{\rho}m + \frac{b}{5}B(t) \right\} \\
&= P \left\{ \sqrt{1-\rho}X(t) < \mu(t) + \frac{b}{5}B(t) \right\} \\
&= P \left\{ \sqrt{1-\rho}X(t) - \frac{b}{5}B(t) < \mu(t) \right\} \tag{A.16}
\end{aligned}$$

where

$$\mu(t) = \frac{\left(bk(t) + (\frac{bt}{5} - 1)\log(I(0)) + (\frac{bt^2}{10} - t)(r - \frac{\sigma_I^2}{2}) \right)}{\sigma_I} + \sqrt{\rho}m$$

Now we note that the LHS of (A.16) is the difference of two normal random variables, and thus also normally distributed. We need to find the mean and

variance of the random variable on the LHS.

$$\begin{aligned}
E[\sqrt{1-\rho}X(t) - \frac{b}{5}B(t)] &= \sqrt{1-\rho}E[X(t)] + \frac{b}{5}E[B(t)] \\
&= 0 \\
Var[\sqrt{1-\rho}X(t) - \frac{b}{5}B(t)] &= Var[\sqrt{1-\rho}X(t)] + Var[\frac{b}{5}B(t)] \\
&\quad (X(t) \text{ and } B(t) \text{ are independent}) \\
&= t(1-\rho) + \frac{b^2}{25} \frac{t^3}{3} \\
&= t(1-\rho) + \frac{b^2 t^3}{75}
\end{aligned}$$

With this result, we let

$$\eta(t) = \sqrt{1-\rho}X(t) - \frac{b}{5}B(t)$$

Then $\eta(t) \sim N(0, t\sqrt{1-\rho} + \frac{b^2 t^3}{75})$. Thus, we get

$$\begin{aligned}
P\{Distress(t)|Default(t)\} &= P\{\eta(t) < \mu(t)\} \\
&= P\left\{\frac{\eta(t)}{\sqrt{t(1-\rho) + \frac{b^2 t^3}{75}}} < \frac{\mu(t)}{\sqrt{t(1-\rho) + \frac{b^2 t^3}{75}}}\right\} \\
&= N\left(\frac{\mu(t)}{\sqrt{t(1-\rho) + \frac{b^2 t^3}{75}}}\right) \tag{A.17}
\end{aligned}$$

If $\rho < 0$, then the derivation is similar and the final result is

$$\mu(t) = \frac{\left(bk(t) + (\frac{bt}{5} - 1)\log(I(0)) + (\frac{bt^2}{10} - t)(r - \frac{\sigma_I^2}{2})\right)}{\sigma_I} - \sqrt{|\rho|}m$$

and

$$P\{Distress(t)|Default(t)\} = N\left(\frac{\mu(t)}{\sqrt{t(1-|\rho|) + \frac{b^2 t^3}{75}}}\right) \tag{A.18}$$

Bibliography

- [1] Acharya, Bharath, and Srinivasan. Does industry-wide distress affect defaulted firms? evidence from creditor recoveries. *Journal of Financial Economics*, 2007.
- [2] V. Acharya and T. Johnson. Insider trading in credit derivatives. *Journal of Financial Economics*, 84, 2007.
- [3] L. Alili, P. Patie, and J. L. Pedersen. Representations of the first hitting time density of an ornstein-uhlenbeck process. *Stochastic Models*, 2005.
- [4] H. Almeida and T. Philippon. The risk-adjusted cost of financial distress. *Journal of Finance*, 2007.
- [5] Altman, Brady, Resti, and Sironi. The link between default and recovery rates: Theory, empirical evidence, and implications. *Journal of Business*, 2005.
- [6] E. Altman and V. Kishore. Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analysts Journal*, 52(6):57–64, 1996.

- [7] Aunon-Nerin, Cossin, Hricko, and Huang. Exploring for the determinants of credit risk in credit default swap transaction data: Is fixed income markets' information sufficient to evaluate credit risk? *FAME Research Papers*, 2002.
- [8] Bakshi, Madan, and Zhang. Understanding the role of recovery in default risk models: Empirical comparisons and implied recovery rates. *Working Paper*, 2006.
- [9] Bharath and Shumway. Forecasting default with the merton distance to default model. *Review of Financial Studies*, 2008.
- [10] F. Black and J. C. Cox. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance*, 1976.
- [11] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 1973.
- [12] Roberto Blanco, Simon Brennan, and Ian W. Marsh. An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. *Journal of Finance*, 60(5):2255–2281, 2005.
- [13] Bolton and Oehmke. Credit default swaps and the empty creditor problem. *Review of Financial Studies*, 2011.
- [14] D. Bongaerts, F. de Jong, and J. Driessen. Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market. *Journal of Finance*, 66, 2011.

- [15] Brigo, Predescu, and Capponi. Credit default swaps liquidity modeling: A survey. *arXiv*, 2010.
- [16] M. Bruche and C. Gonzalez-Aguado. Recovery rates, default probabilities and the credit cycle. *Journal of Banking and Finance*, 34, 2010.
- [17] Long Chen, David A. Lesmond, and Jason Wei. Corporate yield spreads and bond liquidity. *Journal of Finance*, 62(1):119–149, 2007.
- [18] P. Collin-Dufresne and R. Goldstein. Do credit spreads reflect stationary leverage ratios? *Journal of Finance*, 2001.
- [19] P. Collin-Dufresne, R. Goldstein, and J. S. Martin. The determinants of credit spread changes. *Journal of Finance*, 2001.
- [20] Cossin and Hricko. Exploring for the determinants of credit risk in credit default swap transaction data. *FAME research paper*, 2001.
- [21] S. Das, D. Duffie, N. Kapadia, and L. Saita. Common failings: How corporate defaults are correlated. *Journal of Finance*, 2007.
- [22] Sanjiv R. Das and Paul Hanouna. Implied recovery. *Journal of Economic Dynamics and Control*, 33(11):1837 – 1857, 2009.
- [23] D. Duffie, A. Eckner, G. Horel, and L. Saita. Frailty correlated default. *Journal of Finance*, 64(5):2089–2123, 2009.
- [24] D. Duffie and K. J. Singleton. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12(4):687–720, 1999.

- [25] Kenneth French. Data library. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
- [26] Kay Giesecke. Correlated default with incomplete information. *Journal of Banking and Finance*, 28(7):1521–1545, 2004.
- [27] J. M. Harrison. *Brownian Motion and Stochastic Flow Systems*. Wiley, 1985.
- [28] J. Hasbrouck and G. Sofianos. The trades of market makers: An empirical analysis of NYSE specialists. *Journal of Finance*, 48(5):1565–1593, 1993.
- [29] Hilscher, Pollet, and Wilson. Uninformed and inattentive traders in credit default swap markets. *Working Paper*, 2010.
- [30] J. Huang and M. Huang. How much of the corporate-treasury yield spread is due to credit risk. *Working Paper, Stanford University*, 2003.
- [31] Hull, Predescu, and White. The valuation of correlation-depedent credit derivatives using a structural model. *Working Paper*, 2009.
- [32] Prem Jain. The effect on stock price of inclusion in or exclusion from the S&P 500. *Financial Analysts Journal*, 43(1):58–65, 1987.
- [33] R. Jarrow. Default parameter estimation using market prices. *Financial Analysts Journal*, 57(5):75–92, 2001.
- [34] R. Jarrow and S. Turnbull. The intersection of market and credit risk. *Journal of Banking and Finance*, 24, 2000.

- [35] Aditya Kaul, Vikas Mehrotra, and Randall Morek. Demand curves for stocks do slope down: New evidence from an index weights adjustment. *Journal of Finance*, 55(2):893–912, 2000.
- [36] Anh Le. Separating the components of default risk: A derivative-based approach. *Working Paper*, 2007.
- [37] F. Longstaff, S. Mithal, and E. Neis. Corporate yield spread: Default risk or liquidity? new evidence from the credit default swap market. *Journal of Finance*, 60, 2005.
- [38] D. J. Lucas. Default correlation and credit analysis. *Journal of Fixed Income*, March, 1995.
- [39] A. Madhavan and S. Schmidt. An analysis of changes in specialist inventories and quotations. *Journal of Finance*, 48(5):1595–1628, 1993.
- [40] D. Mandan, L. Guntay, and H. Unal. Pricing the risk of recovery in default with absolute priority rule violation. *Journal of Banking and Finance*, 27, 2003.
- [41] Markit. Index methodology for the CDX indices. *www.markit.com*, 2007 and 2012.
- [42] R. Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 1974.
- [43] Pan and Singleton. Default and recovery implicit in the term structure of sovereign cds spreads. *Journal of Finance*, 2008.

- [44] Standard & Poor's. Default, transition and recovery: 2008 annual global corporate default study and rating transitions. *Standard & Poor's Ratings Direct*, 2009.
- [45] J. Qiu and F. Yu. Endogenous liquidity in credit derivatives. *Journal of Financial Economics*, 2012.
- [46] Moody's Investors Service. The u.s. municipal bond rating scale: Mapping to the global rating scale and assigning global scale ratings to municipal obligations. *Moody's Rating Methodology*, 2007.
- [47] Moody's Investors Service. Corporate default and recovery rates, 1920 - 2010. February 28, 2011.
- [48] O. Shachar. Exposing the exposed: Intermediation capacity in the credit default swap market. *working paper*, 2012.
- [49] Shleifer and Vishny. Liquidation values and debt capacity: A market equilibrium approach. *Journal of Finance*, 1992.
- [50] A. Shleifer. Do demand curves for stocks slope down? *Journal of Finance*, 1986.
- [51] S. Shreve. *Stochastic Calculus for Finance II*. Springer, 2004.
- [52] D. Tang and H. Yan. Liquidity and credit default swap spreads. *working paper. University of Hong Kong and University of South Carolina*, 2007.
- [53] C. Zhou. An analysis of default correlations and multiple defaults. *Review of Financial Studies*, 2001.